2012-05-23

Chromatin Dynamics in Pluripotency and Differentiation: A Dissertation

Ozlem Yildirim
University of Massachusetts Medical School

A Dissertation Presented

By

OZLEM YILDIRIM

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2012

Interdisciplinary Graduate Program

Biochemistry and Molecular Pharmacology Department

CHROMATIN DYNAMICS IN PLURIPOTENCY AND DIFFERENTIATION

A Dissertation Presented

By

OZLEM YILDIRIM

Approved as to style and content by:

_____

Oliver J Rando M.D., Ph.D. Thesis Advisor

_____

Alonzo Ross, Ph.D., Chair of Committee

_____

Kami Ahmad, Ph.D., External Member of Committee

_____

Craig Peterson, Ph.D., Member of Committee

_____

Paul Kaufman, Ph.D., Member of Committee

_____

Thomas Fazzio, Ph.D., Member of Committee

_____

Anthony Carruthers, Ph.D., Dean of the Graduate School of Biomedical Sciences

Interdisciplinary Graduate Program

May 23, 2012

DEDICATION

Dedicated to

My beloved family:

Hasan (father), Mefkure (mother) & Alper (brother) Yildirim, Gokce (Sister) & Sarp Kaya (Nephew), Mehmet H. Acar & Mehmet Yildirim (grandfather), Saadet Acar & Rukiye Yildirim (grandmother)

My dear friends:

Dr. Ozlem Topaloglu, Hande Goksal & Ebru Erdem Ortacdag

FOREWORD

Studies detailed in this dissertation have been carried out in Dr. Oliver J. Rando's lab in Biochemistry and Molecular Pharmacology Department at UMASS Medical School. This dissertation consists of 4 chapters and chapter II has been published in a peer-review journal and chapter III is in preparation for submission. I have a separate acknowledgement section at the end of the forth chapter to acknowledge my collaborators and their respective contributions for each chapter in detail.

Foremost, I would like to express my deep gratitude to my advisor Dr. Oliver J Rando for his continued encouragement, motivation and support. He shared with me a lot of his expertise, research insights and critical thinking. I would also like to thank the members of my committee, Dr. Craig Peterson, Dr. Job Dekker, Dr. Paul Kaufman, Dr. Anthony Imbalzano and my chair, Dr. Alonzo Ross. I would like to extend my gratitude to all the past and present members of the Rando lab for creating a productive and friendly work environment. In particular, I would like to thank Dr. Marta Radman-Livaja not only for being collegial and excellent scientist but also for being a wonderful friend and great support. I cannot find enough words to thank Dr. Christopher J Lengner apart from being an excellent mentor and great collaborator; he has also been a dear friend for me. I also would like to thank Faculty in UMASS Medical School, particularly Dr. Thomas Fazzio, who has shared with me a lot of his expertise during our collaboration.

Also, I would like to thank my greatest support and dearest friend Dr. Ozlem Topaloglu and Hande Goksal, I am here because of your endless support. Finally, to all my friends from several different places, as Alfred L Tennyson tells in his poem “I am a part of all that I have met”, thank you for the support and enrichment you brought into my life.

ABSTRACT

Different cell types in multi-cellular organisms heritably maintain different gene expression patterns despite carrying the same genome; a phenomenon termed epigenetics. It is widely believed that the packaging state of the genome, known as chromatin structure, carries epigenetic information.

How chromatin states are inherited and how chromatin structure changes during development, moreover how different epigenomes, such as chromatin and DNA modifications communicate with each other during these processes are important questions. Accordingly, understanding the mechanisms that govern pluripotency and differentiation requires details of chromatin dynamics. The major goal of my doctoral thesis was to understand the genome wide view of chromatin dynamics in embryonic stem cells.

My studies centered on two aspects of chromatin dynamics in mouse embryonic stem cells– localization and function of two antagonistic chromatin regulators and genome-wide histone variant dynamics.

In the first part, we examined the roles of several chromatin regulators whose loss affects the pluripotent state of ES cells. We found that two such regulators, Mbd3 and Brg1, control a large number of genes in ES cells via antagonistic effects on promoter nucleosome occupancy. Moreover, we found that both Mbd3 and Brg1 play key roles in the biology of 5-hydroxymethylcytosine (5hmC), a newly identified DNA modification. Mbd3, which was named by homology to known cytosine methyl binding domains, yet does not bind methylcytosine *in vitro*, co-localized in ES cells with 5hmC. Furthermore, Mbd3

localization was lost in knockdown cells lacking the major 5mC hydroxylase, Tet1. Our results suggest, contrary to current dogma, that 5hmC is more than just an intermediate in cytosine demethylation pathways, that it may regulate genes via the Mbd3/NuRD complex. Finally, we showed that both Mbd3 and Brg1 are themselves required for normal levels of 5hmC *in vivo*, identifying a feedback loop between 5hmC and Mbd3. Together, our results identified a possible effector for 5hmC, thereby suggesting a functional role for this DNA modification. Moreover, Brg1 and Mbd3 can now be added to the growing list of regulators with opposite effects on ES cell gene expression, suggesting that pairs of antagonistic chromatin binding proteins may be a common phenomenon in ES cell transcription regulation (Yildirim et al., Cell 2011).

The second part of my dissertation concerns the dynamics of several histone variants. Seminal studies in the Henikoff lab showed that certain histone variants are replaced throughout the cell cycle, in contrast to the predominant replication-coupled mode of histone assembly. Work in yeast and flies showed that rapid histone turnover occurs at epigenetically-regulated genomic regions, such as chromatin boundary elements or Polycomb/Trithorax binding sites. Notably, promoter regions of actively transcribed genes exhibit rapid turnover, suggesting that histone turnover may have an important role in gene regulation, as higher histone turnover rate would provide higher probability of DNA element exposure and faster erasure of chromatin marks of the replaced histones. In order to extend such studies to a model for pluripotency and differentiation, we developed a system for measuring histone replacement in mouse ES cells. To be

able to carry out turnover experiments in ES cells, we generated stable ES cell lines that can be induced to express epitope-tagged histone variants. Our results confirmed that histone turnover patterns are conserved from yeast to mammals and that turnover profiles are histone variant specific. Murine H3.3 turnover is similar to H3.3 turnover in flies, with peaks at the promoters of highly transcribed genes. MacroH2A2, a variant generally linked to gene repression, had a more complex turnover profile. Surprisingly, we found rapid exchange of macroH2A2 occurring around transcription start sites of a number of highly expressed genes. At poorly expressed genes, on the other hand, macroH2A2 localizes upstream or downstream of transcription start sites and is incorporated slowly, either via slow turnover or via replication-coupled incorporation. Finally, we have used those inducible ES cell lines to generate mice, which will enable future studies on tissue-specific histone replacement *in vivo*.

In summary, my thesis work not only significantly extends our understanding of chromatin regulation in general but also provides a more detailed landscape of chromatin structure and regulation in ES cells. Extending these analyses to differentiating cells and *in vivo* tissue specific dynamics should provide us with a better understanding not only of cell type specific chromatin organization but also improve our ability to program and re-program genomic landscapes *in vitro*.

TABLE OF CONTENTS

LIST OF FIGURES & Tables

LIST OF FIGURES (continued)

CHAPTER 1

Introduction

DNA in eukaryotic cells is organized into chromatin through the packaging of highly conserved basic proteins called histone proteins. Four different histone proteins organize in an octomeric structure; an H3-H4 tetrameric particle in association with two H2A-H2B pairs wraps approximately 147 bp of DNA duplex (**Figure1-1**) (Luger et al., 1997b) and thereby forms the fundamental unit of chromatin called the nucleosome, first described by Kornberg in 1974 (Kornberg, 1974; Kornberg and Thomas, 1974). Nucleosome core particles and the unwrapped connector DNA have been visualized as a "beads on a string" structure with electron microscopy (Olins and Olins, 1974). The linker DNA between nucleosome particles may be associated with the linker histone H1 (Marino-Ramirez et al., 2005; Maxson et al., 1983). These packed nucleosome arrays in association with linker histones and other chromatin proteins allow the higher order organization of chromatin in the nucleus. Both the histone content of nucleosomes and post translational modifications of these histones as well as nucleosome positioning dictate the overall organization of chromatin; which is important for transcription, replication and other DNA-templated processes.

HISTONE GENE REGULATION

In contrast to the majority of eukaryotic genes, which are mostly expressed outside of S phase, the bulk of the core histone protein synthesis starts at the transition from G1 to S phase and peaks simultaneously with DNA replication, although they are expressed at low levels throughout the cell cycle (Maxson et al., 1983; Osley, 1991). In addition to the coordinated induction of

core histone synthesis, the increase in stability of histone mRNA is also an important factor in response to higher histone demand during replication. The inhibition of DNA replication halts histone synthesis and triggers rapid degradation of histone mRNA (Heintz et al., 1983). Moreover, excess histones that are not bound by chaperones or incorporated into chromatin are marked for degradation by Rad53 protein kinase and rapidly degraded in S.cerevisiae (Gunjan and Verreault, 2003). The stoichiometric balance between histone protein content and DNA is very important and therefore, tightly regulated, as nonspecific binding of basic histone proteins to DNA results in insoluble aggregates at physiological ionic strength (Gunjan et al., 2006). Accordingly, gene expression patterns and other mechanisms regulating histone levels in cells are highly conserved (Cho et al., 1998; Cho et al., 2001; Gunjan et al., 2006). Histone synthesis is regulated at several levels, and each organism seems to use different sets of regulatory mechanisms for histone homeostasis (Hentschel and Birnstiel, 1981). It can be regulated at the protein level- through factors that play a role in histone protein storage such as histone chaperones- or at transcriptional and posttranscriptional levels through the regulation of mRNA stability and pre-mRNA processing. Unlike the bulk of Pol II transcribed genes in metazoans, the canonical histone genes do not contain introns and they terminate their transcription with a highly conserved 3’-stem loop structure instead of a poly-A sequence (Osley, 1991). This highly conserved stem loop structure at the 3’end provides an excellent platform for orchestrated synthesis through the regulation of stem loop binding protein (SLBP) (Whitfield et al.,

2000), which is proven to be a crucial protein for histone mRNA processing, translation and degradation (Marzluff et al., 2008). In spite of the similarities in histone proteins and their expression profiles, the transcriptional regulatory mechanisms vary across species (Marino-Ramirez et al., 2006). In all eukaryotes canonical core histone genes and in vertebrates, linker histones as well, are members of multi gene families. In some organisms, histone genes are organized in several tandem repeats spanning five histone genes, while in others, they are mostly dispersed with some clustering (Chaboute et al., 1993; Maxson et al., 1983). Coordinated histone assembly into nucleosomes and DNA synthesis during replication require co-expression of histone genes, which implies mechanisms for transcriptional co-regulation and common features of regulatory elements among different canonical histone genes in an organism. In addition to histone type specific regulatory motifs, the existence of such shared motifs between different histone types has also been demonstrated (Chowdhary et al., 2005; Isogai et al., 2007).

HISTONE MODIFICATIONS

All four nucleosomal histone proteins contain a histone fold domain that shares a common structure and an N-terminal histone tail of variable length (Arents and Moudrianakis, 1995). While the histone fold domain is globular, histone tails are unstructured, mostly protruding out from the nucleosome core particle. Histone proteins are subject to a large number of post translational modifications, especially the residues on amino terminal tails and some on the globular domain, and carboxy terminal tails as well, in the case of H2A and H2B.

These post translational modifications may introduce architectural changes that influence chromatin structure and consequently the regulation of the underlined DNA sequence, as evidenced by the enrichment of different modifications patterns over distinct chromatin domains. Some of the most common modifications are phosphorylation of serine and threonine residues; as well as acetylation or methylation of arginine and lysine residues, in addition to sumoylation or ubiquitylation of select lysine residues on some histone types. These covalent modifications may either influence the nucleosome structure itself by affecting DNA-histone and histone-histone interactions in the core particle or they may have an influence on higher order chromatin structure by affecting histone-histone interactions between neighboring nucleosomes. Moreover, they may provide docking sites for some proteins while they occlude others, thereby defining distinct chromatin domains. Many chromatin proteins and protein complexes are enriched for a number of structural domains that recognize specific modifications. Histone lysine acetylation is recognized by Bromodomain containing proteins. Histone lysine methylation is recognized by the royal family modules which include Chromodomain, Tudor, PWWP, MBT and nonrelated PHD domains. The WD-40 repeat, protein domain that is found in many chromatin associated proteins, is another module implicated in methyl-lysine binding (Kouzarides, 2007; Yap and Zhou, 2006).

Genome wide mapping of single nucleosome histone modifications showed that histone modifications generally occur in broad domains, and they are mostly homogeneous (Liu et al., 2005) with few exceptions such as sharply

localized peaks of H3K4 mono-methylation along with H3K27 acetylation marking the cell type specific functional enhancers (Heintzman et al., 2009). The 5’ ends of gene bodies in transcriptionally active genes are highly enriched for H3K4 di-, tri-methylation, while H3K9 and H3K14 acetylation marks diffuse throughout the gene body. Two-nucleosome hypoacetylation events correspond to 5’ regions of repressed genes whereas large domains of acetylation depletion are characteristic of heterochromatic regions in yeast (Liu et al., 2005). While there are some special acetylation events, such as H3 lysine56 (H3K56) in the histone fold domain, which has been shown to be important for efficient replication dependent nucleosome assembly (Li et al., 2008) and H4K16 acetylation that plays a role in boundary maintenance and prevention of spreading of silencing by Sir proteins (Suka et al., 2002); hyper acetylation of histones is almost always associated with transcriptional activation. It has been shown that acetylation of lysine residues neutralizes its positive charge thereby reduces histone-DNA contacts, thus creates a more open chromatin structure , thought to be favorable for transcription (Dion et al., 2005; Shogren-Knaak et al., 2006). Conversely, the removal of acetylation consistently associates with repression of transcription. Enzymes that reverse the activity of histone acetyltransferases (HATs), are called histone deacetylases (HDACs), and these two opposing enzymes are members of activating and repressive chromatin complexes, respectively (Narlikar et al., 2002). Both HATs and HDACs do not show much substrate specificity especially when compared to lysine methyl transferases that usually modifies a single lysine residue on a specific histone protein. While some of

these methylations, namely on H3K4, H3K36 and H3K79 correlate with transcriptional activity; others particularly H3K9, H3K27 and H4K20 methylations are implicated in transcriptional repression (Barski et al., 2007). Unlike acetylation, ionic interactions are not affected by methylation. Instead, methylated residues provide a binding platform for recruitment of proteins with different methyl binding domains. For example H3K4 tri-methylation at actively transcribed genes recruits the chromatin remodeler CHD1 through its double chromo domain which in return facilitates transcription (Gaspar-Maia et al., 2009) although this mechanism does not apply in yeast (Sims et al., 2005). The core transcriptional machinery can also be recruited to H3K4me3 sites through the affinity of general transcription factor TFIID subunit TAF3 (van Ingen et al., 2008). In yeast, H3K36 tri-methylation deposited behind Pol II recruits the Rpd3S complex to transcribed gene bodies, which suppresses the cryptic transcription initiation events (Carrozza et al., 2005). Members of the enzyme family that methylate arginine residues are called protein arginine methyl transferases (PRMTs). Like lysine methylation, arginine methylation may associate with either transcriptional activation or repression. There are two specific residues known to be ubiquitinylated: one associated with repression and the other with activation, H2AK119 and H2BK120, respectively. A role for histone phosphorylation has been suggested in mitosis, apoptosis and gametogenesis, but histone phosphorylation links to gene expression are not clear. The best known histone phosphorylation event that has been implicated in transcriptional regulation is H3S10 phosphorylation, which recruits chromatin phospho binding protein and

plays a role in NF-kB mediated gene activation (Kouzarides, 2007). (**Figure1-2** & **Figure1-3** summarize some of the well known post translational histone modifications and their modifying enzymes)

HISTONE VARIANTS

Another factor that alters nucleosome structure and greatly influences chromatin domain definition is the replacement of canonical histones with structurally distinct variant histones; even though differences in the primary sequence may be as little as 4-5 amino acids. While canonical histones mostly play a role in monotonous genome packing, variant histones specialize in a range of tasks including chromosome segregation, DNA repair, dosage compensation and germ cell packaging.

Unlike the replication dependent, canonical histones, which have an unusual gene organization, variant histone genes have the regular Pol II gene structure with introns and polyA tails. In addition, rather than being in histone gene clusters as their canonical counterparts, they are dispersed throughout the genome as single copy genes. Histone variants are expressed throughout the cell cycle and incorporated when and where they are needed through the guidance of specific histone chaperones. Phylogenetic analysis of histone proteins demonstrated that canonical histone proteins are one of the most evolutionarily conserved eukaryotic proteins (Malik and Henikoff, 2003). Among them, H4 histone is exceptionally conserved. While all other core histone proteins have non-allelic variants and/or replication independent versions with special

roles, H4 remains the same with hardly any sequence changes among eukaryotic kingdoms (Malik and Henikoff, 2003). All 26 H4 genes, both in mouse and human, encode exactly the same protein. Accordingly, there are no sequence variations between replication coupled and replication independent H4. Replication independent synthesis of invariant H4 and H2B histones are achieved by alternative post-transcriptional processing (Marzluff et al., 2002). In the octameric nucleosome core particle H3 aligns with H2A and H4 with H2B. It was suggested that this pair wise structural alignment may be the reason why both H3 and H2A enriched in different versions but not H4 and H2B (Henikoff and Ahmad, 2005).

H2A VARIANTS

H2A is the core histone with the largest number of variants. While H2Ax and H2Az are present in most eukaryotes, both H2A.Bbd and macroH2A are vertebrate specific H2A variants (Marino-Ramirez et al., 2005) **(Figure1.4)**. H2Ax does not differ much from the canonical H2A aside from an additional C-terminal domain which is crucial for its specific function. Although the mechanism of its incorporation to its target domains is not known very well, it is phosphorylated on Ser139 residue of this conserved C terminal motif in response to DNA damage (Rogakou et al., 1998). A recent study demonstrated that the histone chaperone FACT is responsible for both its incorporation into and dissociation from the nucleosome in human cells. They showed that the phosphorylation of H2Ax by DNA protein kinases facilitates FACT mediated incorporation while PARylation of the FACT subunit Spt16 by PARP1 prevents its disassembly (Heo et al., 2008).

H2Ax phosphorylation quickly spreads around the double strand break sites and plays an important role in the formation of repair foci to which many components of the DNA damage repair machinery are recruited (Fernandez-Capetillo et al., 2004). Independent from its role in genomic stability H2Ax is also crucial for sperm maturation and male fertility due to its role in meiotic sex chromosome inactivation and sex chromosome segregation (Celeste et al., 2002; Fernandez-Capetillo et al., 2003). Furthermore, H2Ax phosphorylation through GABA receptor signaling plays role in S-phase cell cycle control in ES cells, which lack the G1-S cell cycle check point, and regulates embryonic stem cell and neural progenitor cell proliferation (Andang et al., 2008).

Although the structures of H2Az and H2A containing nucleosomes are nearly identical (Suto et al., 2000), H2Az is nevertheless indispensible for many organisms. In mammals, H2Az is encoded by an essential gene and homozygous deletion in mice halts development at the peri-implantation stage (Faast et al., 2001). Similarly, in Drosophila, it has been shown to be imperative for development beyond larval stages (Clarkson et al., 1999). According to biochemical and structural analysis, H2Az stabilizes protein interactions within the same nucleosome and contributes to more stably folded nucleosome particles (Park et al., 2004) while it prevents chromatin condensation by impeding internucleosomal interactions akin to N-terminal histone acetylation (Fan et al., 2002). Its exclusive localization to the transcriptionally active macronucleus in tetrahymena (Stargell et al., 1993) and a positive correlation with transcriptional activity in human cells (Barski et al., 2007; Jin et al., 2009) implicated it in

maintenance of transcriptionally competent chromatin. Besides, it was demonstrated that in human cells RNA Pol II recruitment is potentiated by H2Az enrichment at paused and active promoters, while it's enriched in the body of non-transcribed genes (Hardy et al., 2009). In budding yeast Htz, the yeast orthologue of H2Az, is enriched at the boundary regions that prevent spreading of heterochromatin into euchromatic regions (Meneghini et al., 2003) where hyper dynamic H3 turnover is observed and suggested to contribute to the prevention of heterochromatin spreading (Dion et al., 2007). In contrast with its function in anti-silencing at heterochromatin boundaries, in S. pombe, H2Az works in concert with heterochromatin factors and RNAi machinery to suppress anti-sense RNA (Zofall et al., 2009). On the other hand, Htz assembles in the nucleosomes flanking nucleosome free regions (NFR) and marks both active and inactive genes in euchromatin (Raisner et al., 2005). Similarly, in human cells, transcriptionally active loci are marked with H2Az around the transcription start sites (TSSs) in association with H3.3, a histone variant that is also associated with transcriptional activity. Enrichment for this hybrid nucleosome core particle was also observed at other active regulatory regions (Jin et al., 2009). In addition to its involvement in gene activation there is also evidence linking H2Az to gene repression.. In embryonic stem cells, H2Az was found to be co-localized with polycomb group (PcG) protein Suz12 at developmental regulatory genes that are suppressed in the pluripotent state. And this co-localization at target promoters is interdependent as depletion of one reduced the enrichment of the other. Furthermore, depletion of H2Az leads to activation of these genes, underscoring

its role in repression (Creyghton et al., 2008). Another strong support for its role in gene inactivation was the demonstration of the role of the extended acidic patch of H2Az in recruitment of silencing protein HP1α which in return promotes higher order chromatin formation (Fan et al., 2004). In any case, it is not yet clear how H2Az positively or negatively influence transcription.

Another H2A variant, H2ABbd is one of the most diverged versions of H2A. It has 48% primary sequence similarity to canonical H2A; it lacks the typical C-terminus tail and has a truncated docking domain, which tethers the H2A-H2B dimer to the H3-H4 tetramer (Malik and Henikoff, 2003). Therefore, these relatively short H2A variant containing nucleosomes only wrap 130bp of DNA (Doyen et al., 2006b). It was first recognized by its exclusion from the inactive X (Xi) chromosome in female mammalian cells and co-localization with H4K12Ac, which is indicative of euchromatin function (Chadwick and Willard, 2001b). Reconstitution studies suggested that H2ABbd-containing nucleosomes are looser and less stable than conventional H2A-containing nucleosomes. *In vivo* support to these studies came from FRAP experiments, showing that the fluorescence recovery of H2ABbd was faster than that of H2A (Gautier et al., 2004). Nevertheless, H2ABbd containing nucleosomes are shown to be resistant to SWI/SNF remodeling. Unlike macroH2A containing nucleosomes, H2ABbd nucleosomes cannot be remodeled by either SWI/SNF or ACT even when remodeling is facilitated by the histone chaperone nucleolin (Angelov et al., 2006). This resistance to remodeling is due to the differences in the histone fold domain (Doyen et al., 2006b). Although it is found to be expressed in different

female tissues, a recent study demonstrated its presence in advance mouse spermiogenesis and mature human sperm. And it is suggested that H2ABbd may facilitate histone displacement by protamins during sperm maturation (Ishibashi et al., 2010), although it is also conceivable that it may be important for the function of nucleosomaly organized regions during and after fertilization.

Like H2A.Bbd, macroH2A is a vertebrate specific H2A variant (Marino-Ramirez et al., 2005). H2A variants exhibit the most pronounced primary structure heterogeneity in their C-terminal domain. MacroH2A (also known as H2AFY) has a very distinct C-terminal domain that includes a small stretch of basic amino acids and a putative leucine zipper motif. While the N-terminal domain of the protein shares 65% amino acid sequence similarity with the canonical H2A, the C-terminal domain (also known as the macro domain) is almost twice as big as the canonical H2A and has no relation to other histones (Chadwick and Willard, 2001a). Interestingly, the C-terminal tail of H2A variants has the potential to interfere with H1 binding (Suto et al., 2000). The crystal structure of the macroH2A containing nucleosome core particle suggests that some of the structural differences, when compared to the canonical nucleosome structure, render it potentially more stable and less flexible (Chakravarthy et al., 2005a). Biochemical analysis, showing high ionic strength and resistance to dissociation in lower salt concentrations (Abbott et al., 2005), further strengthens that notion.

Despite being localized on different chromosomes- macroH2A1 on 13 and macroH2A2 (also known as H2AFY2) on 10- the two genes share a similar gene

structure with conserved exon-intron boundaries except for a spliced variant site in macroH2A1. MacroH2A1 encodes two spliced variants, macroH2A1.1 or macroH2A1.2 and both are conserved in mammals. MacroH2A2 is 80% identical to macroH2A1.2 and it is 98% conserved between mice and human (Chadwick and Willard, 2001a; Costanzi and Pehrson, 2001; Zhang et al., 2005). Even though they are preferentially expressed at different levels in different cell types in the same tissue, the two proteins seem to localize to the same genomic targets. MacroH2A1.1 differs from the other two forms in several ways, the functional consequences of which are presently not clear. In contrast to the splice variant macroH2a1.1, both macroH2a1.2 and macroH2a2 co-localize to inactive X chromosomes (Xi) in a distinct chromatin structure named the macro chromatin body (MCB) (Costanzi and Pehrson, 1998). They are also found in centrosomes (Chadwick and Willard, 2002) and senescence associated foci (Zhang et al., 2005). Interestingly, MacroH2A1.1 expression was shown be restricted to terminally differentiated cells (Pehrson et al., 1997) and anti correlated with proliferation markers (Sporn et al., 2009). Additionally, macroH2A1.1 is the only form that can bind to the SirT-1 metabolite, O-acetyl-ADP-ribose (AAR) (Kustatscher et al., 2005) and the PARP1 metabolite PAR (poly ADP-ribose) (Timinszky et al., 2009). The interaction of the macro domain of macroH2A1.2 with PARP1, although possibly non-specific (Timinszky et al., 2009), inhibits PARP1 enzymatic activity (Nusinow et al., 2007).

MacroH2A enrichment of the Xi chromosome in female mammalian cells was interpreted as an evidence for its role in silencing (Costanzi and Pehrson,

1998), even though the association of macroH2a with the Xi appears to occur after random inactivation of the additional X chromosome (Chadwick and Willard, 2001a). Moreover, macroH2a localization to Xi depends on the non coding RNA XIST but it is not required for the inactive state. When XIST is conditionally deleted, gene silencing is maintained despite the loss of macroH2a on Xi (Csankovszki et al., 1999).

More evidence for its role in transcriptional repression stems from *in vitro* findings. The C-terminal 'macro' domain of macroH2A1.2 was shown to repress transcriptional initiation by almost 10 fold, mostly by preventing p300 HAT activity and chromatin remodeling (Doyen et al., 2006a). Additionally, another study demonstrated that macroH2A incorporation into nucleosome prevents the binding of the transcription factor NF-kB to nucleosomal DNA, while also it renders the nucleosome resistant to the SWI/SNF and ACF remodeling (Angelov et al., 2003). Later, the same group demonstrated that the nuclear protein nucleolin, a FACT like histone chaperone that has a role in transcriptional activation, promotes the SWI/SNF remodeling of macroH2A containing nucleosomes (Angelov et al., 2006).

As evidenced by its localization to autosomes, macroH2a's functions go beyond its role in MCB formation in Xi. Additionally, organisms, such as chicken, that do not undergo X inactivation, still express this H2A variant (Abbott et al., 2005). ChIP-Chip experiments demonstrated the co-localization of macroH2A1.2 and macroH2A2 with the polycomb repressive complex 2 (PRC2) at developmentally regulated genes in the testicular cancer cell line NT2.

Consistent with its suggested role in gene repression, macroH2A occupancy was reduced upon induction of some of its target genes. Furthermore, when macroH2A levels were reduced by RNAi, transcriptional activation of these genes was stronger (Buschbeck et al., 2009). Moreover, the Kraus group showed that macroH2A1 occupies large domains in autosomal chromosomes and mostly associates with transcriptional repression. In the positively regulated subset of its target genes, macroH2A was localized to coding regions. Interestingly, those positively regulated genes are involved in signal mediated transcription such as serum starvation, and RNAi depletion of macroH2A causes the loss of the serum starvation response, suggesting a direct role for macroH2A in transcriptional activation of these genes (Gamble et al., 2010). Similar results were found in mouse embryonic stem cells. During the retinoic acid induced neuronal differentiation of ES cells, transcriptional activation of retinoic acid targets and some other differentiation related genes depends on macroH2A1.2 occupancy of these genes. On the other hand, repression of core pluripotency factors upon induction of the differentiation program was less efficient in the absence of macroH2A1.2. Therefore, authors propose a role for macroH2A in maintaining the balance between self-renewal and differentiation of both adult and embryonic stem cells (Creppe et al., 2012).

Meanwhile, several recent studies implicated macroH2A in human disease. In one study macroH2A expression was shown to be positively correlated with a good prognosis of lung cancer due to the strong negative correlation between tumor cell proliferation and macroH2A protein levels (Sporn

et al., 2009), while in another study, RNAi depletion of macroH2a in dormant melanoma cells was shown to render them metastatic (Kapoor et al., 2010). Additionally, macroH2A mislocalization through ATRX deficiency has been proposed to contribute to the alpha-thalassemia phenotype of the ATRX syndrome (Ratnakumar et al., 2012). Elevated expression of macroH2A1 in Huntington Disease (HD) is another example of histone dysregulation in human disease. The mouse model for HD confirmed the progressive up-regulation of macroH2A1 expression with disease progression. Even though there isn't a suggested mechanistic role for macroH2A1 in HD pathology both the prognosis and macroH2A1 levels are responsive to HDAC inhibitor treatment (Hu et al., 2011). Interestingly, the C-terminal non histone region (macro domain) was shown to bind HDAC1 and 2 (Chakravarthy et al., 2005a).

HISTONE H3 VARIANTS

In addition to two versions of replication dependent H3 core histones H3.1 and H3.2, there are three replication independent H3 variants: a centromere specific H3 CenH3, a sperm specific H3t (Witt et al., 1996) and H3.3, which is expressed throughout the cell cycle and maintained during differentiation (Ahmad and Henikoff, 2002; Wu and Gilbert, 1981) **(Figure1.5)**.

CenH3 is a centromere specific H3 variant with 50 to 60% similarity to the canonical H3 in the histone fold domain and no sequence similarity in the N-terminal tail which can be highly variable across different species (Malik and Henikoff, 2003). CenH3 is another essential histone variant and the homozygous

deletion of CenpA, the mammalian centromere specific H3, in mice causes post-implantation lethality (Howman et al., 2000). Centromeres are chromosomal domains that anchor kinetechores, which attach mitotic spindle fibers and allow faithful chromosome segregation. Kinetechore formation at centromeres depends on a special chromatin structure and especially the presence of CenH3 (Collins et al., 2005). It has been suggested that it is centromere specific histones, rather than their underlying DNA sequences, that specify centromeres, since they are a more constant feature of centromers as opposed to the highly variable satellite DNA sequences found in those regions (Henikoff et al., 2004). Besides, human neocentromeres that are devoid of satellite DNA sequences can still assemble into centromeric structure. Furthermore, mislocalized CenH3 is subject to ubiquitin mediated rapid degradation (Collins et al., 2004). This high fidelity of CenH3 in annotating and maintaining specific chromatin domain have histone variants considered as prime candidates for carriers of epigenetic memory (Henikoff et al., 2004).

H3.3 is another highly conserved H3 variant whose expression is not restricted to S phase. In human, mouse and flies H3.3 is expressed from 2 separate genes, H3.3A and H3.3B. Both genes encode the exact same protein sequence from mRNA sequences with distinct promoter elements (Frank et al., 2003) and un-translated regions (Bramlage et al., 1997), suggesting differential transcriptional and post-transcriptional regulation. H3.3 is associated with transcriptionally active genes and active regulatory elements (Ahmad and Henikoff, 2002; Goldberg et al., 2010), and has been involved in transcriptional

elongation and the replacement of nucleosomes evicted by the transcriptional machinery (Wirbelauer et al., 2005). In line with these findings, H3.3 is more enriched for transcriptional activity associated chromatin marks (Hake et al., 2006; McKittrick et al., 2004; Sutcliffe et al., 2009). Its association with transcriptional activation is further supported by the observation that H3.3 nucleosomes are depleted of heterochromatin binding protein1 (HP1) (Loyola et al., 2006) and they are refractory to linker histone H1 (Braunschweig et al., 2009). In addition, another role for its involvement in transcriptional activation is indicated by the high H3.3 turnover rates at active regulatory elements as this hyper dynamic histone exchange possibly contributes to the accessibility of underlying DNA elements by transcription factors (Henikoff, 2008). Furthermore, the interference with canonical H3 incorporation in the early mouse embryo by depletion of subunits of the replication coupled chromatin assembly pathway results in increased H3.3 incorporation and consequently impaired heterochromatin formation (Akiyama et al., 2011). Surprisingly, H3.3 was shown to be enriched in the telomeric heterochromatin of murine ES cells and this enrichment correlates with normal telomere function and integrity (Goldberg et al., 2010; Lewis et al., 2010). H3.3 may replace H3.1 in a neutral way as well, in non-proliferating terminally differentiated cells after mitotic exit (Pina and Suau, 1987) and during DNA repair outside of S phase (Zhang et al., 2007) where H3.3 compensate for H3.1 to maintain consistent nucleosome density. Hence, there appear to be two different mechanisms for H3.3 assembly into nucleosomes, one for processes that replace canonical H3 outside of S phase and the other for

producing functionally distinct transcriptionally permissive chromatin domains. The distinct distribution of H3.3 in the genome (Drane et al., 2010; Goldberg et al., 2010; Mito et al., 2005) and its importance for certain cell types (Orsi et al., 2009; Sakai et al., 2009; Torres-Padilla et al., 2007) strongly suggest that H3.3 is more than just a surrogate for H3.1 and emphasizes its crucial role in normal chromatin regulation. However, H3.3 deletion in drosophila has been shown to be viable and its depletion was compensated by up-regulation of canonical H3 genes. While H3.3 depletion is tolerable in terms of viability it was indispensible for fertility as homozygous mutants were invariably sterile (Sakai et al., 2009), suggesting a role for H3.3 in transitioning from nucleosomal chromatin to protamins during sperm maturation or during chromatin reorganization at fertilization. In agreement with this idea, the H3.3 specific chaperone HIRA (Bonnefoy et al., 2005; Loppin et al., 2005) and the chromatin remodeler CHD1 (Konev et al., 2007) are critical for paternal chromatin rearrangement during fertilization and early embryo development. Male fertility was also found to be severely compromised in mice with H3.3 defects; H3.3A gene disruption not only resulted in reduced fertility but also caused partial neonatal lethality in homozygous mutants (Couldrey et al., 1999), despite existence of an intact H3.3B gene which produces the same protein product. In mammals, nucleosomes evicted during chromatin reorganization in meiotic sex chromosome inactivation (MSCI) and silencing of unsynapsed autosomal chromatin (MSUC) are replaced exclusively with H3.3 containing nucleosomes. The authors of the study argue that such nucleosome turnover and incorporation

of newly synthesized H3.3 provide a blank slate for the redecoration of nucleosomes with specific histone modifications necessary for normal chromatin formation in the germ line (van der Heijden et al., 2007).

HISTONE CHAPERONES & CHROMATIN REMODELERS

The preferential incorporation of histone variants with distinct functions to specific chromatin domains in eukaryotic genomes suggests that there are specialized deposition mechanisms for targeting of histone variants to particular loci. How different histone variants are targeted to specific chromatin domains and what are the structural and functional consequences of variant incorporation to distinct loci are important questions. Several studies aimed to answer these questions and confirmed the role of different histone chaperones and chromatin remodelers in the differential localization of histone variants.

Histone chaperones are defined as factors that associate with histones to escort them to their destination and facilitate their assembly into chromatin. It is suggested that the characteristic property of a histone chaperone is histone transfer but not necessarily the ability to function as a histone assembly factor (De Koning et al., 2007) **(Table1.1)**. Histone chaperones play important roles in preventing nonspecific association of highly basic histone proteins with nucleic acids by binding free histones and escorting them to their deposition sites. Replication coupled histone deposition onto nascent DNA is governed by the three subunit protein called Chromatin Assembly Factor 1 (CAF-1). The canonical H3s- H3.1 and H3.2-associate with CAF-1 that travels behind the

replication fork with the proliferating cell nuclear antigen (PCNA) and they are incorporated into chromatin in a replication coupled manner (Shibahara and Stillman, 1999). While in yeast the sole H3 variant associates with both Caf1 and HirA, in vertebrate HIRA binds specifically to the replication independent H3 variant H3.3 (Tagami et al., 2004). The amino acid substitutions in the histone fold domain in H3.3 allow this selective association with replication independent chromatin assembly factors (Ahmad and Henikoff, 2002). On the other hand,. both chaperones receive the newly synthesized H3-H4 dimers from the same histone chaperone Asf1 (Tagami et al., 2004), which also mediates transcription coupled eviction as well as deposition of H3-H4 dimers (Schwabish and Struhl, 2006). Recent studies show that HIRA is not the only chaperone responsible for H3.3 deposition. In fact, various chaperones facilitate the differential targeting of H3.3 throughout mammalian chromatin. While HIRA is mostly involved in H3.3 deposition at promoter regions of active genes and in transcription coupled deposition, the Daxx-Atrx complex is required for H3.3 localization at pericentric heterochromatin and telomeres (Drane et al., 2010; Lewis et al., 2010; Wong et al., 2010). DEK is another novel H3 chaperone which has been shown to have H3.3 specificity *in vitro* and is potentially responsible for H3.3 deposition at regulatory elements (Sawatsubashi et al., 2010).

Because the eukaryotic genome is organized into chromatin, accessibility of nucleosomal DNA, as well as regulation of DNA templated processes depend on the cooperation between several factors including sequence specific DNA binding proteins, histone and DNA modifications, histone variants, repressors,

activator and mediator complexes of the general transcriptional machinery working in coordination with chromatin regulatory factors such as histone modifying enzymes, histone chaperones and ATP dependent chromatin remodelers. All known chromatin remodelers contain an ATP hydrolyzing Snf2p-related catalytic core which is homologous to the large helicase super family 2(SF2) and they are usually part of a large protein complex such as the SWI/SNF chromatin remodeling complex (Flaus and Owen-Hughes, 2004). Like many Snf2p-related proteins, the SWI/SNF complex works directly on nucleosomes and unravels the octamer in association with histone chaperones (Clapier and Cairns, 2009; Cote et al., 1994). *In vivo* findings show that transcription through nucleosomes requires dimer displacement. Accordingly chromatin remodeling complexes at sites of RNA Pol2 activity were shown to be associated with the histone chaperone FACT (Belotserkovskaya et al., 2003). Biochemical analysis also implicated similar cooperation between the histone chaperone Nap1 and the ATP dependent chromatin remodeling factor RSC *in vitro* (Lorch et al., 2006). In addition, the chromatin remodeling machinery may re-position nucleosomes as a secondary result of DNA distortion they induced, since several ATP dependent chromatin remodelers are capable of DNA-translocation activity (Saha et al., 2002).

Remodeling enzymes are usually recruited to their target location by other regulatory proteins, such as sequence specific transcription factors, transcriptional co-activators and general transcription machinery components or by transcriptional repressors (Goldmark et al., 2000; Peterson and Workman,

2000). Another way of targeting involves direct binding to histone modifications or to histone tails through specific recognition domains, such as the recruitment of SWI/SNF to specific acetylation patterns via the bromodomain of Snf2 protein (Hassan et al., 2002) or the recruitment of ISWI to histone tails or to linker histones through its SANT domain (Boyer et al., 2004). Likewise, the CHD family of remodelers can be recruited by methylated histone H3 tails through its tandem chromodomains (Flanagan et al., 2005). Like SWI/SNF, INO80 family utilizes its helicase-SANT-associated (HSA) domain which allows its association with actin related protein subunits (Szerlong et al., 2008) such as Arp4 and Arp8 that can bind to histones and may also have a histone chaperone function (Watanabe and Peterson, 2010). Therefore, chromatin remodeling complexes work in coordination with histone modifying enzymes and enable recruitment of transcriptional regulators to target loci (Cosma et al., 1999). Different subunit compositions within remodeling complexes define them both functionally and structurally and provide them with specialized functions. For instance, embryonic stem cell specific SWI/SNF-like remodeling complex, esBAF is characterized by several subunits which allow its functional role in pluripotency by co-occupaying and regulating the function of core pluripotency genes (Lessard and Crabtree, 2010). In the mean time, different families of chromatin remodeling complexes are decorated for executing distinct tasks. While SWI/SNF complex activity usually disrupts initially equally spaced nucleosomal arrays to promote activation; the ISWI family plays a central role in nucleosome assembly and nucleosome positioning to produce ordered nucleosomal arrays, mostly to promote repression

(Goldmark et al., 2000). However, there are examples of uncharacteristic activity for each family, such as SWI/SNF promoting repression by facilitating repressor binding (Martens and Winston, 2002) and ISWI assisting transcriptional elongation by organizing nucleosomal arrays in the gene body (Morillon et al., 2003). The two members of the INO80 family, INO80 and SWR-C, also have distinct functions. While INO80 uses ATP hydrolysis to mobilize and relocate nucleosomes on DNA and has several functions including replication fork progression, transcriptional activation and DNA repair, SWR1 is incapable of DNA translocation reaction. On the other hand, it has a unique ability to catalyze dimer exchange; specifically replacing canonical H2A-H2B dimers with variant H2Az-H2B dimers in a unidirectional manner (Luk et al., 2010). A recent biochemical study in support with genome wide ChIP analysis showed that the INO80 complex can counteract SWR1 activity and specifically replace H2Az-H2B dimers with H2A-H2B. Accordingly, INO80 deletion causes abnormal spreading of H2Az as well as genome instability in yeast (Papamichos-Chronakis et al., 2011; Watanabe and Peterson, 2010). The fourth chromatin remodeling family is the CHD family, characterized by two N-terminal tandem chromodomains and a SNF2 like ATPase domain. The CHD family has 9 different members with distinct functions. In mammals CHD1 is thought to be involved in transcriptional activation because of its association with pre-initiation factors and has also been implicated in facilitating transcriptional elongation as well as splicing (Sims et al., 2007). CHD3 and CHD4, also known as Mi-2α and Mi-2β, respectively, lack the DNA binding domain but harbor a PHD (Plant Homology Domain), which is an

alternative methyl lysine binding domain. CHD3/4 is the member of a large protein complex called NURD (nucleosome remodeling and deacetylation complex). HDAC1, HDAC2, MTA1, MTA2, MTA3, RbAp48, RbAp46, MBD2 or MBD3 are the other subunits of NURD. Although NURD could be classified as a repressor according to its subunit content, it is linked to both gene repression and activation (McDonel et al., 2009). The chromatin remodeling activity of NURD was suggested to facilitate access to acetylated nucleosomes as the HDAC activity of the complex was shown to be stimulated by ATP hydrolysis (Tong et al., 1998). A limited number of studies suggests that NURD may be recruited to its targets through transcription factors (Kim et al., 1999); although, it has also been shown that MBD2 containing NURD but not MBD3 containing NURD can be recruited to methylated DNA (Le Guezennec et al., 2006).

DNA METHYLATION & HYDROXYMETHYLATION

DNA methylation is another epigenetic modification that influences cell type specific gene expression profiles and enables the inheritance of these expression profiles during cell division. Three enzymes with DNA methyl transferase activity are responsible for the establishment and maintenance of mammalian DNA methylation patterns. In mammals, cytosines, that are followed by guanine (CpG dinucleotide), receive the methylation mark. Initial methylation patterns are established by two catalytically active de novo DNA methyl transferases Dnmt3a and Dnmt3b; and during imprinting in association with the catalytically inactive Dnmt3l (Hata et al., 2002). Maintenance of the existing DNA methylation patterns during replication is provided by the methyl transferase

Dnmt1 which interacts with PCNA in replication foci and preferentially methylates hemi-methylated CpG dinucleotides (Chuang et al., 1997). All three catalytically active DNA methylation enzymes are essential for viability as mice, deficient for those genes, are either embryonically lethal or post natal lethal (Wu and Zhang, 2010). DNA methylation patterns, while relatively stable in somatic cells, are erased and reestablished in the zygote and the pre-implantation embryo respectively (Wu and Zhang, 2010). This “do over” of methylation patterns is critical for the establishment of pluripotency and normal development (Feng et al., 2010). It is therefore not surprising that abnormal DNA methylation patterns are frequent in tumor cells (Baylin and Jones, 2011).

CpG dinucleotides are distributed disproportionately in mammalian genomes. Low density CpG dinucleotides are scattered throughout the genome and are mostly methylated. High GC content regions of approximately one thousand base pairs in size, with discrete CpG dinucleotide rich clusters are called CpG islands (CGIs). CGIs usually mark transcription initiation sites and are frequently not methylated in mammalian genomes except in the intergenic “orphan” CGIs in which CpG methylation is more frequent (Deaton and Bird, 2011). CGI containing promoters include housekeeping genes, developmental regulators and tissue specific genes; covering around 70% of annotated promoters in vertebrate genomes (Saxonov et al., 2006). These promoters have a distinct chromatin structure and transcriptional regulation. The transcriptionally permissive nature of CGI promoters is demonstrated by genome wide transcriptome analysis both with RNA Pol II ChIP-Seq (Guenther et al., 2007)

and global nuclear run-on experiments (Core and Lis, 2008). In agreement with its transcriptionally permissive nature, CGI chromatin is suggested to be intrinsically accessible and does not require ATP dependent chromatin remodeling activity (Deaton and Bird, 2011). Accordingly, it is also enriched for transcriptional activity associated histone modifications such as hyperacetylation of H3 and H4, as well as the exclusion of linker histone H1 (Tazi and Bird, 1990). Moreover, genome wide studies revealed that H3K4me3 is a signature histone modification of CGI chromatin, independently of the promoter's activity state in human and mouse cells (Guenther et al., 2007; Mikkelsen et al., 2008). Similarly, in mouse embryonic stem cells, unmethylated high density CpG dinucleotides are associated with housekeeping genes or developmentally important genes, while methylated low density CpG regions correspond to tissue specific gene promoters. CpG methylation levels and histone modifications are highly correlated: housekeeping genes with a high CpG dinucleotide content are highly expressed and enriched for the transcriptional activity associated histone modification H3K4me3. On the other hand developmentally regulated genes carry both the activation and repression associated histone modifications H3K4me3 and H3K27me3, respectively, and they are transcriptionally inactive (Meissner et al., 2008).

Methylation of CGIs provides stable silencing of relevant genes during differentiation and is necessary for normal development (Mohn et al., 2008). Silencing via DNA methylation is thought to be either by inhibition of transcription factor binding due to methylated DNA itself or through the recruitment of methyl

binding domain proteins, which mediates the establishment of repressive chromatin (Deal et al., 2010; Klose and Bird, 2006). In line with this idea, all known interacting partners of methyl binding domain (MBD) containing proteins are transcriptional repressors and their association with MBD family members further supports their silencing activity (Bogdanovic and Veenstra, 2009). However, as evident from the X chromosome inactivation, DNA methylation is a later event in silencing; it occurs after the establishment of repressive chromatin modifications such as histone hypoacetylation as well as transcriptional silencing. Rather than playing a role in the initiation of the silencing, DNA methylation provides the stability of silencing (Csankovszki et al., 2001). Hence, genomic regions in need of long term silencing such as imprinted genes, transposons and Xi tend to be methylated.

Although DNA methylation is a faithfully heritable, stable epigenetic modification, there are occasions where demethylation occurs; either by passive dilution of methylation in the absence of Dnmt1 during replication or by an active enzymatic process. Existence of genome wide (for example the demethylation of the paternal genome during embryonic development) and gene specific DNA demethylation events suggested the existence of active DNA demethylation mechanisms. Several pathways were proposed for active DNA demethylation reaction including base excision repair after DNA glycolysation or deamination, as well as nucleotide excision mechanisms in coordination with the GADD45A DNA damage response function. These proposed mechanisms, however, fail to explain rapid genome wide demethylation activity (Wu and Zhang, 2010). On the

other hand oxidative demethylation mechanisms and TET (ten-eleven translocation) family proteins appear to be better candidates for global demethylation enzymes. Tet1, the first identified member of the Tet protein family, was shown to be responsible for the generation of 5-hydroxymethylcytosine (5hmC) from 5meC both *in vitro* and vivo (Tahiliani et al., 2009). Later, the other two members of the Tet family; Tet2 and Tet3 enzymes were also shown to be capable of hydrolyzing 5meC to 5hmC (Ito et al., 2010).

Recent genome wide studies demonstrated that Tet1 localizes at CpG rich promoters along with Polycomb proteins (Wu et al., 2011c), suggesting that it is another factor that defines CGI chromatin in murine ES cells. On the other hand, 5hmC is found both in promoter regions and gene bodies of ES cells (Wu et al., 2011a). The signature histone modification H3K4me3, found at CGI promoters, is catalyzed by the MLL and SET1 protein families, which are both recruited to unmethylated CGI promoters by a CXXC domain, found in the proteins themselves or their interacting partners (Illingworth et al., 2008). Tet1, which also contains a CXXC domain, was proposed to be responsible for the depletion of methylation at CGIs by catalyzing the oxidation of 5meC to 5hmC (Wu et al., 2012). In line with this idea, Tet1 depletion in ES cells leads to increased DNA methylation at many Tet1 target promoters (Williams et al., 2011; Wu et al., 2011a).

Despite the general idea that hydroxymethylation is primarily an intermediate stage in cytosine demethylation (Bhutani et al., 2011), the relatively persistent steady-state levels of 5hmC observed in several contexts (for

example, during global “demethylation” of paternal DNA after fertilization (Inoue and Zhang, 2011b; Iqbal et al., 2011) suggest otherwise. Its enrichment, especially in certain cell types such as purkinje neurons (Kriaucionis and Heintz, 2009) and embryonic stem cells, suggests that hydroxymethylation may also serve as another DNA modification. It may facilitate the formation of distinct chromatin domains by recruiting or excluding certain factors and serve a specific regulatory function. In line with this idea, MBD family proteins, with intact methyl binding domains which allow their recruitment to methylated DNA, are incapable of binding hydroxymethylated DNA *in vitro* (Jin et al., 2010). While the Tet1 KO mice are viable and fertile (Dawlaty et al., 2011), Tet1 and Tet2 depletion in ES cells results in differentiation defects with abnormal activation of the extra-embryonic gene expression program (Koh et al., 2011). Interestingly, depletion of the Mbd3 protein- which cannot bind methylated DNA (Hendrich and Bird, 1998) in ES cells leads to similar defects as the ones observed in Tet1 KD in ES cells, exhibiting increased expression of some trophectoderm markers (Kaji et al., 2006) (Zhu et al., 2009).

PLURIPOTENT CHROMATIN & EPIGENETIC MEMORY

The zygote and the early blastomeres until the 8 cell stage have the greatest lineage potential called totipotency, which indicates the ability to generate both embryonic and extra-embryonic lineages. By the time that the embryo reaches the 16 cell stage, the first differentiation event occurs when totipotent blastomeres give rise to either outer cells or inner cells of the late morula stage, precursor cells of extra-embryonic (trophectoderm) or embryonic

lineages (inner cell mass), respectively (Surani et al., 2007). Embryonic stem (ES) cells are derived from the inner cell mass (ICM) of pre-implantation embryo. ES cells (ICM *in vivo*) have two characteristic features: pluripotency and self renewal. The first defines their potential to adopt different cellular programs and give rise to all the cells of the three primary germ layers of the embryo. The second describes their capacity to inherit this permissive state through an unlimited number of apparently symmetrical divisions without differentiating (Loebel et al., 2003; Meissner et al., 2009; Nishikawa et al., 2007). Moreover, they may be induced to acquire new transcription programs *in vitro* that are reminiscent of the developmental programs *in vivo* and differentiate into specified cells while they gradually lose their developmental potential and permissive state for alternative fates. Therefore, they are an ideal model for studying chromatin regulation and its role in mammalian development as well as epigenetic mechanisms involved in genomic and cellular plasticity.

One of the most curious questions about an ES cell is “what makes an ES cell capable of assuming any cellular type in the organism?” Recently, many studies aimed to answer this question by focusing on ES cell specific transcriptional networks (Boyer et al., 2005; Wang et al., 2006). Data obtained from those studies identified proteins that control chromatin structure and accessibility. These findings not only emphasize the role of epigenetic mechanisms in the maintenance of cell type specific transcription programs, but also suggest their relevance in the maintenance of pluripotency. Accordingly, genetic ablation of factors ranging from Polycomb to the TIP60 complex to

components of the NURD complex and others have all been shown to have defects in either the maintenance of the pluripotent state, or defects in differentiation (Boyer et al., 2006; Creyghton et al., 2008; Fazzio et al., 2008b; Kaji et al., 2006).

Pluripotent chromatin is transcriptionally more permissive and this open chromatin structure is actively maintained by chromatin remodeling factors, as the depletion of Chd1 in ES cells leads to the increase in heterochromatic foci and features that associate with it including HP1 and H3K9me3 (Gaspar-Maia et al., 2009). Besides, undifferentiated ES cells express considerably higher amounts of general transcription factors and chromatin remodelers; when compared to neural precursor cells (Efroni et al., 2008). Furthermore, microscopy studies revealed distinct features of heterochromatin in pluripotency. While larger and fewer domains of heterochromatic regions show a looser structure in ES cells, upon differentiation heterochromatin become condensed and more abundant (Meshorer and Misteli, 2006). The flexible arrangement of pluripotent chromatin has become more obvious in FRAP (Fluorescence Recovery after Photobleaching) experiments. In differentiated cells, the inner core of nucleosomes is known to be stable, while H2A-H2B dimers on the surface of the nucleosome exchange more readily (Kimura and Cook, 2001). Conversely, FRAP studies in ES cells showed that inner core histones have a “hyperdynamic” nature. Their high fluorescence recovery suggests a high histone turnover rate (Meshorer et al., 2006). Compellingly, fluorescence recovery rates substantially decrease upon differentiation to neural precursor cells suggesting a role for

histone turnover in the maintenance of genomic plasticity. Correspondingly, ES cell chromatin is more enriched for transcriptional activity associated histone modifications (Efroni et al., 2008) and for the replacement histone H3.3 (Hake et al., 2006; McKittrick et al., 2004; Sutcliffe et al., 2009).

The single cell zygote, gives rise to more than 200 different cell types in the mammalian body. In an organism, all these cell types essentially carry the same genome but still manage to establish and maintain their cellular identity. One of the major distinctions between differentiated cell types is that different populations of cells express specific subsets of their genome. Accordingly, establishment and maintenance of cellular identity depends on regulatory mechanisms that control specific transcriptional profiles, relevant for that particular cell type. Such specific transcriptional programs are maintained and established by specific gene activation and silencing processes when and where they are needed (Orlando, 2003). Moreover, these differential gene expression patterns are heritable and different in different types of cells with identical DNA sequence. Such maintenance of specific phenotypes by non-genetic means is called epigenetic inheritance. It has been suggested that much of the heritable epigenetic information is carried by chromatin structure. Nucleosomal positioning, protein composition of the histone octamer (with canonical or variant histones) and covalent modification states of the histone proteins shape chromatin structure and are therefore potential information carriers (Henikoff et al., 2004; Rando, 2007). Accordingly, chromatin regulatory mechanisms are involved in the

maintenance and inheritance of cellular memory as well as genomic plasticity (Theise and Wilmut, 2003).

Genetic screens in Drosophila for mutants with developmental defects identified Polycomb (PcG) and Trithorax (TrxG) group proteins as factors required for memory of differentiated states (Ringrose and Paro, 2007). These proteins were later shown to act as chromatin modifying proteins – Polycomb repressors and some members of Trithorax activators both chemically modify histone proteins and alter chromatin compaction, acting as central players in epigenetic programming (Boyer et al., 2006). Thus, the interplay between PcG/TrxG proteins and chromatin structure regulates the genome organization which affects gene expression and epigenetic inheritance. Recently, PcG/TrxG proteins have been implicated as key factors in the control of chromatin structure in embryonic stem cells (Ang et al., 2011; Boyer et al., 2006). In the pluripotent state, genes controlling developmental pathways, especially lineage specific transcription factors, are repressed by Polycomb group proteins and are preferentially de-repressed upon differentiation (Bracken et al., 2006; Lee et al., 2006a). Interestingly, histone methylation patterns typical of both PcG and TrxG-repressive and active marks, respectively- co-exist at promoters of developmentally important transcription factors in embryonic stem cells but not in their differentiated counterparts (Bernstein et al., 2006). Moreover, these so called ‘bivalent domains’; of the chromatin of genes that are poised for activation but kept repressed, resolve upon differentiation. Genes with bivalent domains get either activated and loose the repressive mark- methyl-H3K27- or become

permanently repressed and lose their active mark- methyl-H3K4 (Bernstein et al., 2006). Thus, this contentious state of chromatin may have a role in ES cell plasticity by priming genes involved in developmental regulation and providing a flexible arrangement in pluripotent genome organization. Accordingly, bivalent genes are depleted of DNA methylation which provides more stable repression, while many of those become methylated and transcriptionally silenced upon induction of differentiation. Furthermore, depletion of de novo DNA methyl transferases compromises the ES cells differentiation potential but their self-renewal properties (Trowbridge and Orkin, 2010), Conversely, the depletion of the maintenance DNA methyl transferase Dnmt1 leads to premature differentiation of progenitor cells and loss of self-renewal (Sen et al., 2010).

Although tissue specific transcription factors play important roles in defining cellular identity, chromatin is essential for maintaining and stabilizing it. While a defined set of transcription factors is sufficient to reprogram differentiated cells into an embryonic stem cell state (Takahashi and Yamanaka, 2006; Wernig et al., 2007) or to another differentiated cell state (Zhou et al., 2008), reprogrammed cells show preference to re-differentiate into their original cell type (Hu et al., 2010) and they may carry residual DNA methylation signatures of their tissue of origin (Kim et al., 2010b; Kim et al., 2011). This suggests that erasing epigenetic memory and resetting chromatin is not easy to accomplish. Moreover, chromatin factors are essential not only for experimental reprogramming but also for natural reprogramming during zygotic genome activation and primordial germ cell (PGC) development. Chromatin regulators such as DOT1L, Suv39h1 (Onder

et al., 2012), and Brg1 (Singhal et al., 2010) affects the efficiency and kinetics of cell reprogramming while, Brg1 (Bultman et al., 2006) is essential for switching from the zygotic to the embryonic program, and Nap1 mediated histone replacement is important for the erasure of epigenetic memory during PGC reprogramming (Hajkova et al., 2008). These findings indicate that chromatin factors play roles in both maintenance and disruption of epigenetic states and successive transmission of these differential expression profiles through several generations relies on epigenetic memory. Accordingly, a more in depth knowledge of chromatin dynamics and chromatin regulatory mechanisms is required for a better understanding of the mechanisms that govern pluripotency and differentiation.

Therefore I focused on different aspects of pluripotent chromatin not only to have a better insight in chromatin dynamics and epigenetic mechanisms that are involved in regulation of stemness but also to understand the details of this permissive chromatin organization. In chapter two we explore the details of interactions between different chromatin remodeling factors with respect to DNA modifications and their interacting partners. While in chapter three, we focused on the histone turnover phenomenon and present genome wide organization and dynamics of different histone variants both in embryonic stem cells and their committed counterparts.

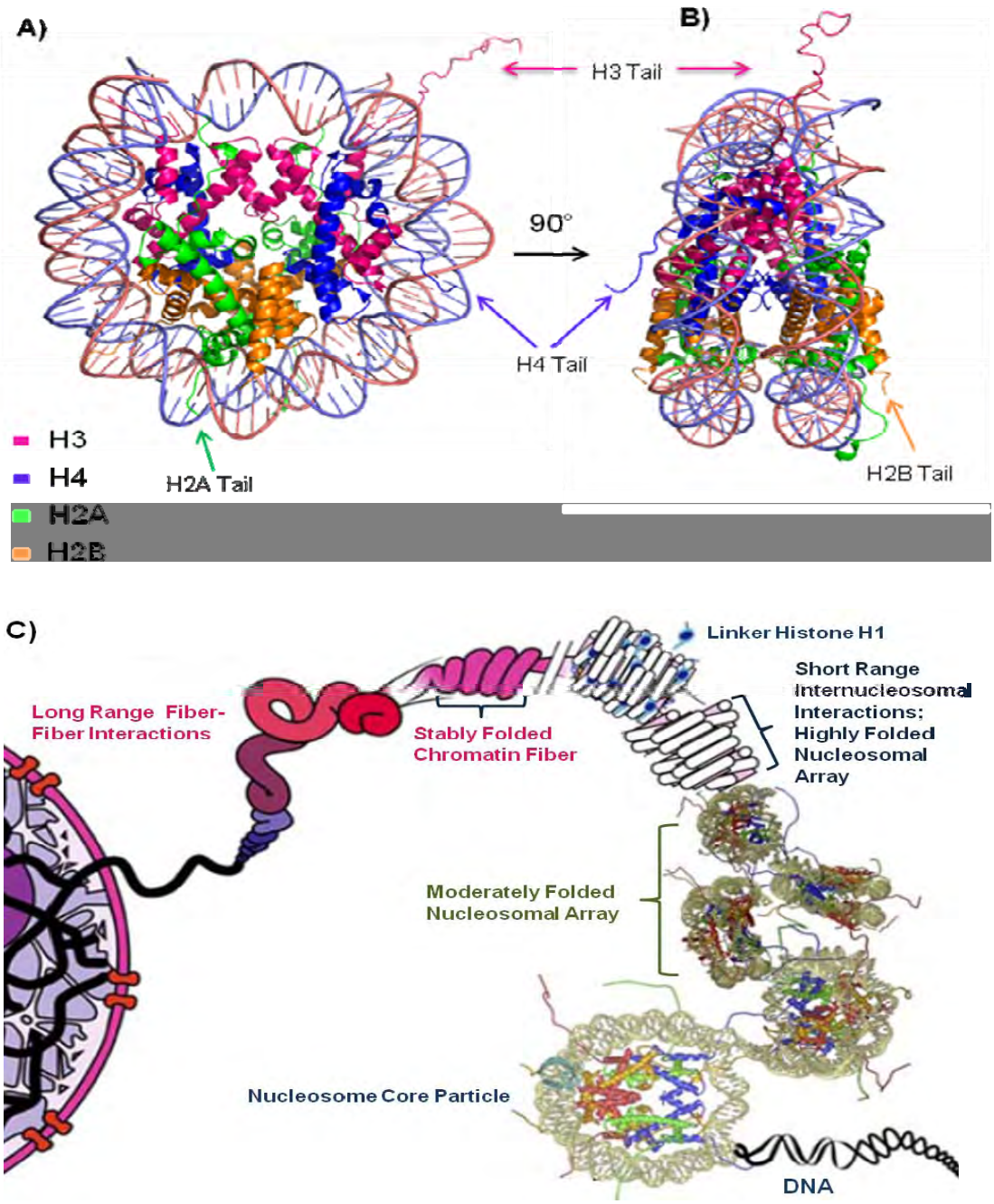

A)
B)
H3 Tail
90°
H4 Tail
H3
H4
H2A
H2B
H2A Tail
H2B Tail
C)
Linker Histone H1
Short Range Internucleosomal Interactions; Highly Folded Nucleosomal Array
Long Range Fiber-Fiber Interactions
Stably Folded Chromatin Fiber
Moderately Folded Nucleosomal Array
Nucleosome Core Particle
DNA

Figure 1-1

Figure 1-1. Packaging of eukaryotic DNA into chromatin

A) Front view of nucleosome core particle (protein data base 1AOI) (Luger et al., 1997a). H3, H4, H2A and H2B are shown in pink, blue, green and orange respectively. Histone octamer, containing a pair of each histone proteins, is wrapped by approximately 147 base pairs of DNA in 1.67 superhelical turns. Unstructured histone tails are viewed protruding from the core particle.

B) Side view of the same structure obtained by 90º rotation

C) Adapted from Chakravarthy S. et al., (Chakravarthy et al., 2005b) DNA in eukaryotes organized into chromatin via the histone protein complex, nucleosome core particle. Internucleosomal interactions provides moderate to high levels of nucleosomal array folding, while highly folded nucleosomal arrays' long range interactions allows higher order chromatin fiber organization.

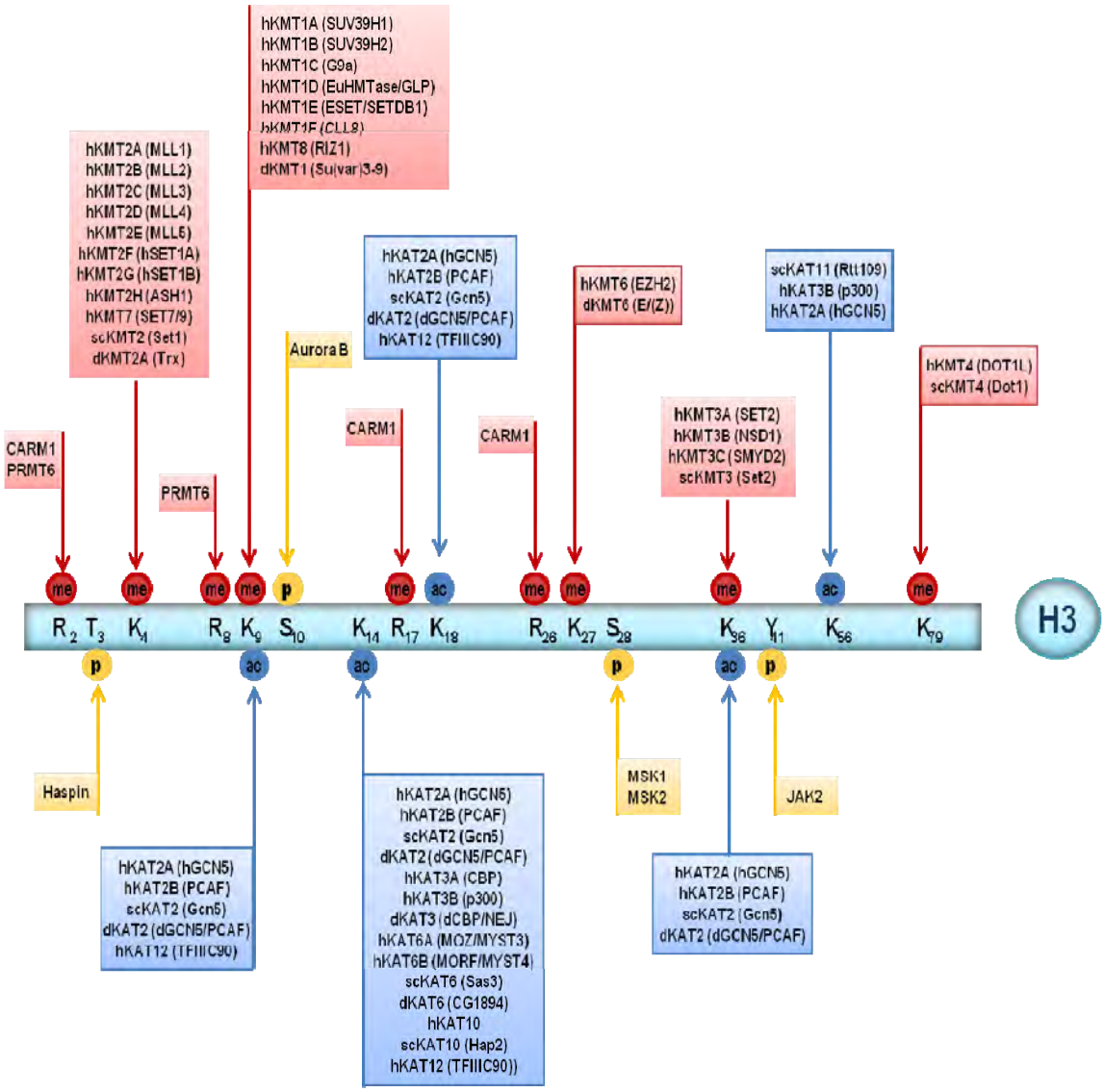

hKMT1A (SUV39H1)
hKMT1B (SUV39H2)
hKMT1C (G9a)
hKMT1D (EuHMTase/GLP)
hKMT1E (ESET/SETDB1)
hKMT1F (CLL8)
hKMT8 (RIZ1)
dKMT1 (Su(var)3-9)
hKMT2A (MLL1)
hKMT2B (MLL2)
hKMT2C (MLL3)
hKMT2D (MLL4)
hKMT2E (MLL5)
hKMT2F (hSET1A)
hKMT2G (hSET1B)
hKMT2H (ASH1)
hKMT7 (SET7/9)
scKMT2 (Set1)
dKMT2A (Trx)
Aurora B
hKAT2A (hGCN5)
hKAT2B (PCAF)
scKAT2 (Gcn5)
dKAT2 (dGCN5/PCAF)
hKAT12 (TFIIIC90)
hKMT6 (EZH2)
dKMT6 (E/(Z))
scKAT11 (Rtt109)
hKAT3B (p300)
hKAT2A (hGCN5)
hKMT4 (DOT1L)
scKMT4 (Dot1)
CARM1
PRMT6
PRMT6
CARM1
CARM1
hKMT3A (SET2)
hKMT3B (NSD1)
hKMT3C (SMYD2)
scKMT3 (Set2)
me
me
me
me
p
me
ac
me
me
me
ac
me
R2 T3 K4 R8 K9 S10 K14 R17 K18 R26 K27 S28 K36 Y41 K56 K79
H3
p
ac
ac
p
ac
p
Haspin
hKAT2A (hGCN5)
hKAT2B (PCAF)
scKAT2 (Gcn5)
dKAT2 (dGCN5/PCAF)
hKAT12 (TFIIIC90)
hKAT2A (hGCN5)
hKAT2B (PCAF)
scKAT2 (Gcn5)
dKAT2 (dGCN5/PCAF)
hKAT3A (CBP)
hKAT3B (p300)
dKAT3 (dCBP/NEJ)
hKAT6A (MOZ/MYST3)
hKAT6B (MORF/MYST4)
scKAT6 (Sas3)
dKAT6 (CG1894)
hKAT10
scKAT10 (Hap2)
hKAT12 (TFIIIC90))
MSK1
MSK2
JAK2
hKAT2A (hGCN5)
hKAT2B (PCAF)
scKAT2 (Gcn5)
dKAT2 (dGCN5/PCAF)

Figure 1-2

Figure 1-2. Histone H3 modifications and modifying enzymes.

A large number of histone posttranslational modifications occur at histone H3 N-terminal tails but there are specific sites that are modified within the globular histone core such as H3K56, H3Y41 and H3K79; acetylated, phosphorylated and methylated respectively. Numerous lysine residues are acetylated in H3 tails including, H3K9, H3K14, H3K18 and H3K36. While all known H3 acetylation marks associate with transcriptional activation, lysine and arginine methylation can either associate with activation as in the cases of H3K4me3 and H3R17me or be involved in repression as in the cases of H3K27me3, H3K9me3 and H3R2me.

(Phosphorylation, acetylation and methylation events shown by yellow circles marked with P, blue circles marked with ac and red circles marked with me respectively. Enzymes responsible from the modification of specific residues are mentioned in the boxes with corresponding colors.)

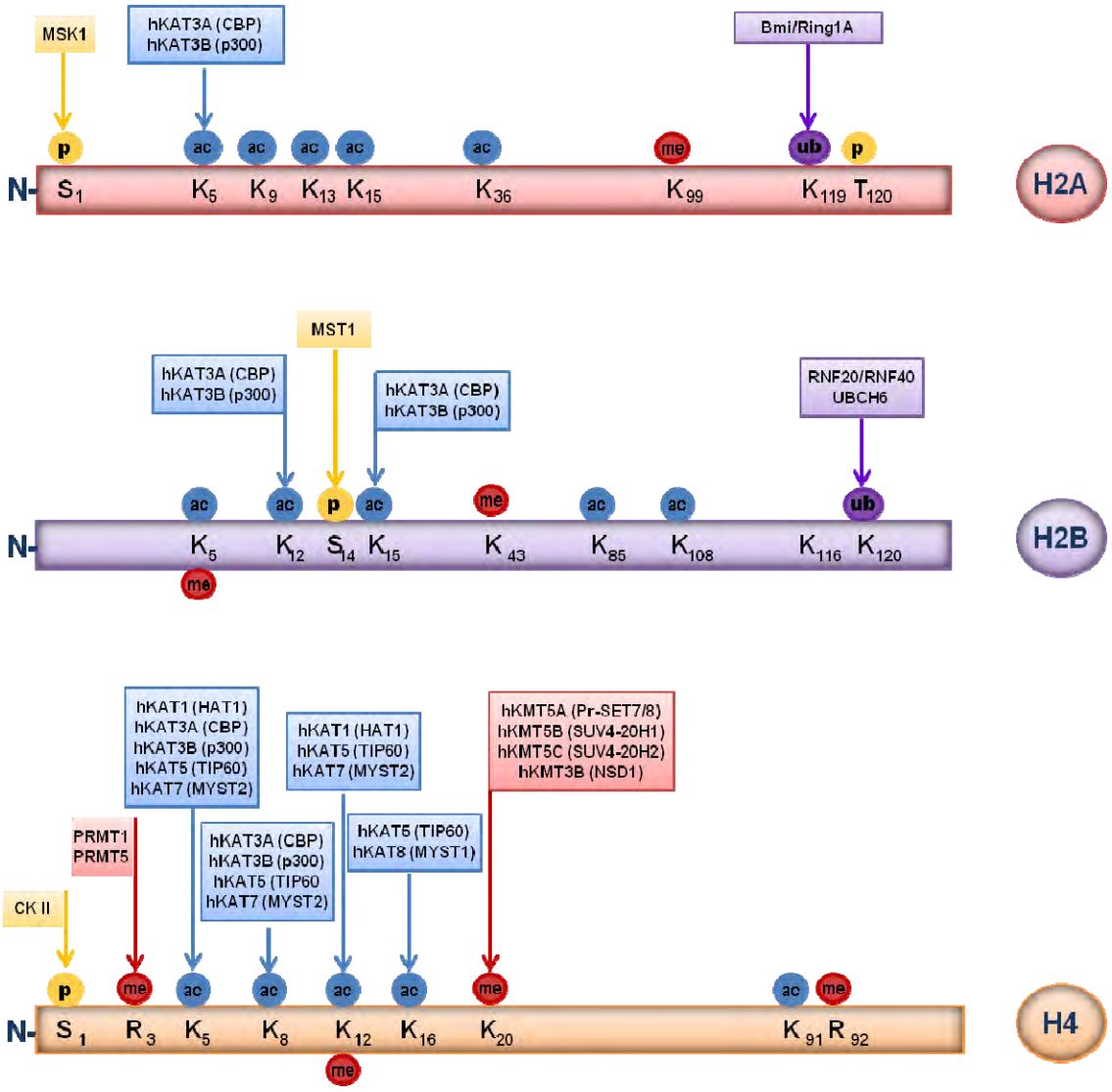

MSK1
hKAT3A (CBP)
hKAT3B (p300)
Bmi/Ring1A
p
ac
me
ub
N-
S1 K5 K9 K13 K15 K36 K99 K119 T120
H2A
MST1
hKAT3A (CBP)
hKAT3B (p300)
RNF20/RNF40
UBCH6
S14
K5 K12 K15 K43 K85 K108 K116 K120
H2B
hKAT1 (HAT1)
hKAT3A (CBP)
hKAT3B (p300)
hKAT5 (TIP60)
hKAT7 (MYST2)
hKAT8 (MYST1)
hKMT5A (Pr-SET7/8)
hKMT5B (SUV4-20H1)
hKMT5C (SUV4-20H2)
hKMT3B (NSD1)
PRMT1
PRMT5
CK II
S1 R3 K5 K8 K12 K16 K20 K91 R92
H4

Figure 1-3

Figure 1-3. Histone H2A, H2B and H4 modifications and modifying enzymes. H2A/H4S1 phosphorylation is enhanced in mitotic chromatin during chromatin condensation. Two ubiquitylation events one in H2AK119 and the other in H2B120 associate with repression and activation respectively. H2AK119 is ubiquitylated by Bmi/Ring1A , a protein found in the polycomb complex. H2B is ubiquitylated by RNF20/RNF40 and UbcH6 in humans and promotes facilitate transcriptional elongation. Two specific methylation events occur in H4 tails. While H4R3 methylation associates with hyperacetylation via recruitment of p300 and transcriptional activation, H3K20 methylation plays a role in the formation and maintenance of heterochromatin as well as DNA repair.

(Phosphorylation, acetylation, methylation and ubiquitylation events shown by yellow circles marked with P, blue circles marked with ac, red circles marked with me and purple circles marked with ub respectively. Enzymes responsible for the modification of specific residues are mentioned in the boxes with corresponding colors.)

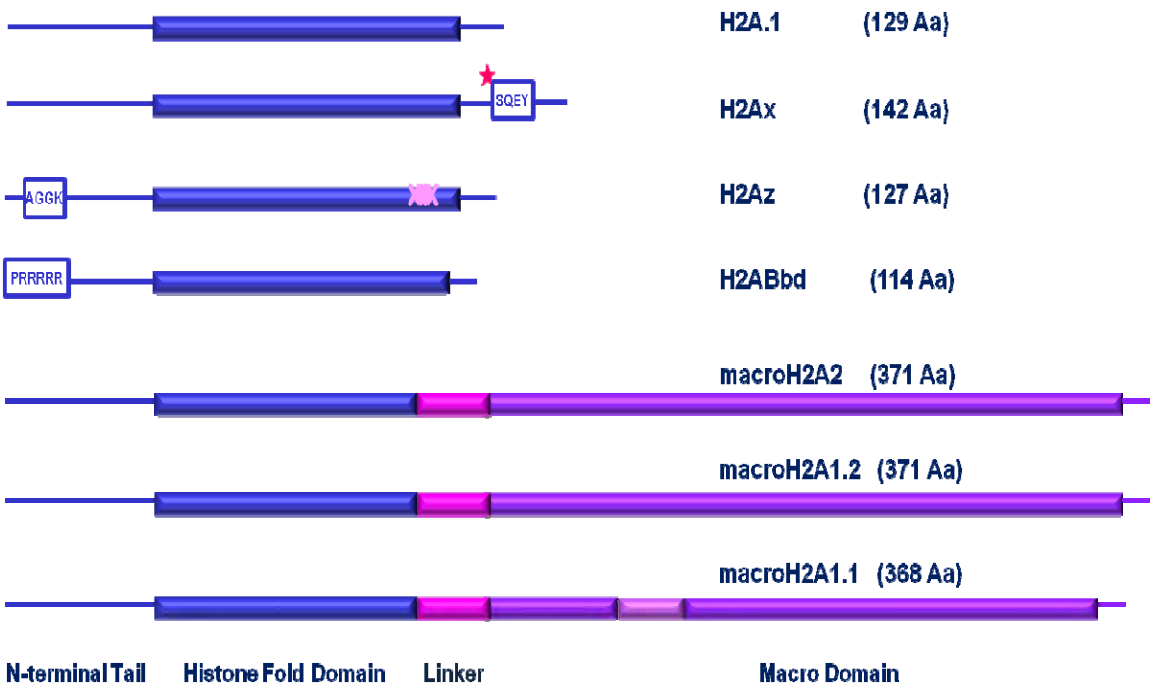

H2A.1 (129 Aa)
SQEY
H2Ax (142 Aa)
AGGK
H2Az (127 Aa)
PRRRRR
H2ABbd (114 Aa)
macroH2A2 (371 Aa)
macroH2A1.2 (371 Aa)
macroH2A1.1 (368 Aa)
N-terminal Tail
Histone Fold Domain
Linker
Macro Domain

Figure 1-4

Figure 1-4. Vertebrate Histone H2A variants.

Histone H2A is the most diverse histone type, while H2Ax and H2Az are universal H2A variants, present in most eukaryotes, macroH2A and H2ABbd are vertebrate specific. H2Ax has a similar histone fold domain to canonical H2A but has a signature consensus sequence (SQEY) at its C-terminal which gets phosphorylated at S139 residue upon DNA damage (star signifies the phosphorylation event). The essential H2A variant H2Az is highly conserved between species (95% conservation between mammalian and urchin), and it is 59% conserved between mammalian H2Az and H2A1 (Hatch and Bonner, 1988)**.** The acidic patch in canonical H2A (DEE) extended by one residue (to DEELD); this extended acidic patch (signified with pink patch on histone fold domain) has been shown to be important for its deposition and function in chromatin. H2ABbd is only 48% identical to canonical H2A and it is particularly divergent with its shorter docking domain. MacroH2A is almost 3 times bigger than canonical H2A due to its non histone macro domain and linker region. But its N-terminal third is 64% identical to full length canonical H2A (Pehrson and Fried, 1992). MacroH2A1 has two spliced variants, MacroH2A1.1 and 1.2. MacroH2A2 is most similar to MacroH2A1.2 with 80% identical amino acid sequence except for the alternative splice site. Alternatively spliced exon 6 of MacroH2A1.1 is shown in pink in the macro domain.

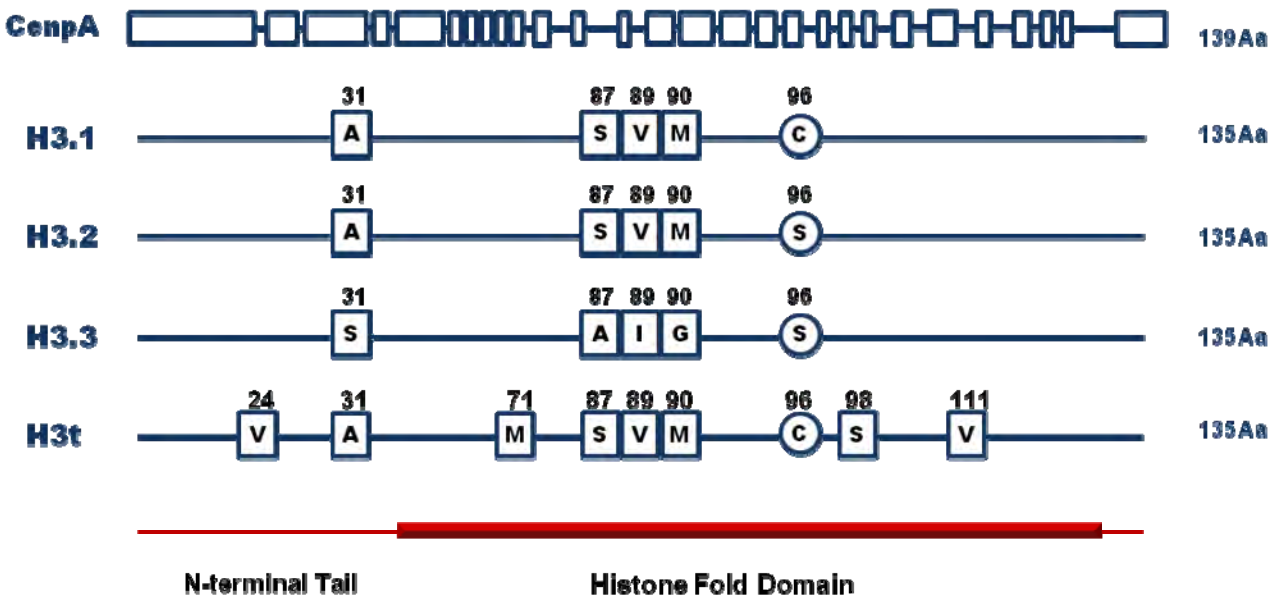

CenpA
139Aa
H3.1
31 87 89 90 96
A S V M C
135Aa
H3.2
31 87 89 90 96
A S V M S
135Aa
H3.3
31 87 89 90 96
S A I G S
135Aa
H3t
24 31 71 87 89 90 96 98 111
V A M S V M C S V
135Aa
N-terminal Tail
Histone Fold Domain

Figure 1-5

Figure 1-5. Mammalian Histone H3 variants.

There are two isotypes of replication dependent H3s; H3.1 and H3.2. These canonical H3s differ by one amino acid at their 96th residue cytosine/ serine respectively. Replication independent H3.3 and testis specific H3t only differs by 5 amino acids (at different residues) from canonical H3.1. Whereas CENPA, mammalian centromeric histone H3, has only 60% identity at histone fold domain of canonical H3s and has no sequence similarity in N-terminal tail.

Histone Chaperones		Histone Selectivity	Main Functions	Complex Association
Asf1a & Asf1b		**H3.1**-H4 & **H3.3**-H4	Histone donor for CAF-1 and HIRA	
tNasp & sNasp		H3-H4	H3-H4 storage	
Spt6		H3-H4	Transcription initiation & elongation	
Daxx/Atrx		**H3.3**-H4	DNA synthesis independent deposition at pericentric heterochromatin & telomeres	
DEK		**H3.3**-H4	DNA synthesis independent deposition at regulatory elements	
HJURP		**CENPA**-H4	Centromeric chromatin formation	
Nucleophosmin (NPM1, 2 & 3)		H2A-H2B	Cytosolic-nuclear transport, replication, transcription	
Chz1 (*Sc.*); HIRIP3? (Hs)		**H2AZ**-H2B	H2AZ incorporation by SWR1	
Nap1 Nap1related proteins: Nap1L2 (*Mm*),SET/TAF1b(Hs), CINAP (*Hs, Mm*)		H2A-H2B	Cytosolic-nuclear transport Transcription Replication	
Nucleolin (*Hs, Ms*)		H2A-H2B	Transcription elongation; assist chromatin remodeling	
CAF-1 complex	p150	**H3.1**-H4	DNA synthesis dependent deposition: Replication & Repair	
	p60			
	RbAp48			
HIRA complex	HIRA	**H3.3**-H4	DNA synthesis independent deposition	
	Ubinuclein			
	Cabin1			
FACT complex	Spt16	H2A-H2B	Transcription elongation	
	SSRP1	H3-H4		
Rsf-1		H3-H4	Assist remodeling complex	RSF
Arp4		ND	Assist remodeling complex & HAT	INO80, HAT Tip60
Arp7, Arp9		ND	Assist remodeling complex	SWI/SNF & RSC BAF/PBAF&BAP
Arp8		H3-H4	Assist remodeling complex	INO80
Acf1		H2A-H2B & H3-H4	Assist remodeling complex	ACF/CHRAC
RbAp48		**H3.1**-H4	DNA synthesis dependent deposition	CAF-1 complex, Remodeling Complex, HAT, HDAC, PC
		H3.3-H4	DNA synthesis independent deposition	
		CENPA-H4	Centromeric chromatin maintenance	

Table1-1

Table 1-1. Histone Chaperones and their main functions.

Histone chaperones, their histone type specificity and main functions are summarized (Adapted from De Koning et al., 2007) (De Koning et al., 2007).

CHAPTER 2

Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells

The following work is reprinted from the Cell article of the same name published as:

Ozlem Yildirim[1], Ruowang Li[2], Jui-Hung Hung[2], Poshen B. Chen[3], Xianjun Dong[2], Ly-sha Ee[3], Zhiping Weng[1,2], Oliver J. Rando[1†] and Thomas G. Fazzio[3,4†]

Cell, 23 December 2011: Vol. 147, Issue 7, pp. 1498–1510.

ABSTRACT

Numerous chromatin regulators are required for embryonic stem (ES) cell self-renewal and pluripotency, but few have been studied in detail. Here, we examine the roles of several chromatin regulators whose loss affects the pluripotent state of ES cells. We find that Mbd3 and Brg1 antagonistically regulate a common set of genes by regulating promoter nucleosome occupancy. Furthermore, both Mbd3 and Brg1 play key roles in the biology of 5-hydroxymethylcytosine (5hmC): Mbd3 co-localizes with Tet1 and 5hmC *in vivo*, Mbd3 knockdown preferentially affects expression of 5hmC-marked genes, Mbd3 localization is Tet1-dependent, and Mbd3 preferentially binds to 5hmC relative to 5-methylcytosine *in vitro*. Finally, both Mbd3 and Brg1 are themselves required for normal levels of 5hmC *in vivo*. Together, our results identify an effector for 5hmC, and reveal that control of gene expression by antagonistic chromatin regulators is a surprisingly common regulatory strategy in ES cells.

In human and mouse embryonic stem (ES) cells, a number of transcriptional regulators are essential to maintain the pluripotent state. Several transcription factors are required for pluripotency, including Oct4, Sox2 and Nanog, which function as "master regulators" of the ES cell transcriptional network (Young, 2011). Along with sequence-specific transcription factors, many chromatin regulators also play essential roles in ES cell gene regulation, self-renewal, and differentiation.

Brg1 and Mbd3 have opposing effects on expression of shared target genes

To better understand chromatin regulation of the ES cell transcriptional regulatory network, we sought to identify the transcriptional targets and functional relationships among six chromatin regulators with important roles in ES cell self-renewal and pluripotency: Tip60, p400, Suz12 (involved in H3K27 methylation), Ash2l (involved in H3K4 methylation), Mbd3 and Brg1. We measured genome-wide mRNA changes upon knockdown (KD) of each factor alone or in all pair-wise combinations (to identify synergistic or epistatic relationships) in murine ES cells (**Table S2-1)**. Consistent with previous data (Fazzio et al., 2008b), KD of Tip60 or p400, alone or in combinations with other factors, exhibited similar changes in gene expression (**Figures 2-1A, S2-1A**), while Brg1 and Mbd3 KD profiles were poorly correlated with these gene expression effects. Principal component analysis (PCA) confirmed that KD of Tip60 or p400 elicited the strongest changes in gene expression (**Figure 2-1B**).

Surprisingly, Brg1 KD and Mbd3 KD affected the same principal component as one another but in opposite directions, suggesting that these two factors oppositely regulate expression of a common set of targets. Indeed, the overlap of genes misregulated upon Brg1 or Mbd3 KD was highly significant ($p < 2.2e-16$). As genes that were similarly up- or down-regulated in both Mbd3 KD and Brg1 KD cells were also nonspecifically affected by other KDs (**Figure S2-1B**), we therefore we focused on the large group of genes *upregulated* upon Mbd3 KD and *downregulated* upon Brg1 KD (**Figure 2-1C**). We validated several of these common Brg1/Mbd3 targets by quantitative RT-PCR (RTqPCR) (**Figure S2-1C**), and found that expression of these genes was also altered in independent Mbd3 or Brg1 KDs using non-overlapping esiRNAs (**Figure S2-1D**) and in KDs of additional subunits of either the Mbd3/NURD or BAF complexes (**Figure S2-1E**). We next searched for epistatic effects between Mbd3 and Brg1. Double KD of Brg1 and Mbd3 resulted in a more wild type mRNA profile than either single KD, indicating that these proteins play antagonistic roles at their common target genes (**Figures 2-1C and S2-1C**).

What features do Mbd3/Brg1 target genes share in common? In normal ES cells, Mbd3/Brg1 target genes are expressed at moderate to high levels (**Figure 2-1D**). Shared targets between these complexes were enriched for a number of functional categories, several of which relate to cellular adhesion and signaling **(Table S2-2)**. Among the common targets of Brg1 and Mbd3, we noted genes encoding signaling molecules with important functions in ES cell self-renewal or differentiation, including *Wnt3a* and *Tgfb1* **(Figure S2-1C)**. These

findings suggest that Mbd3/NURD and BAF complexes function in opposition to fine-tune the expression of a set of genes required for ES cell viability or self-renewal.

Mbd3 binds just downstream of the TSS

Are the joint Brg1/Mbd3-regulated genes direct targets of these complexes? While Brg1 has been mapped genome-wide in ES cells (Ho et al., 2009a), Mbd3 localization is currently unknown. We therefore carried out ChIP-Seq for Mbd3 in murine ES cells, finding that Mbd3 localized largely to promoters (**Figures 2-2A-C, S2-2, S2-3A,B**), with only weak localization to enhancers (**Figures S2-4A,B**). Genes upregulated upon Mbd3 KD were associated with somewhat higher levels of Mbd3 than were unaffected genes (**Figure S2-3C**), although, as is commonly observed with chromatin regulators (Fan et al., 2005; Jiang et al., 2011; Rando and Chang, 2009; Wu et al., 2011b; Wu et al., 2011c), only a small subset of Mbd3-bound loci are transcriptionally affected by Mbd3 loss (**Figure 2-2A**).

Comparing Mbd3 maps to prior maps of Brg1 binding (Ho et al., 2009a), we found significant overlap between the binding sites of the two factors (**Figure 2-2B, S2-4C**). We validated Brg1 and Mbd3 binding at several common target promoters in ES cells by qPCR (**Figure S2-3B**). Interestingly, Mbd3 binding typically occurred ~100-200 bp downstream of the average peak of Brg1 (**Figure 2-2C**), suggesting that these complexes bind in an oriented fashion with respect to the direction of transcription (see below). Furthermore, both Mbd3 and Brg1

were significantly associated with their antagonistically-regulated target genes (**Figure 2-2D, Figure S2-4D**).

How are Brg1 and Mbd3 recruited to common promoters? We first asked whether Brg1 and Mbd3 physically interact *in vivo* using IP-Westerns. Immunoprecipitation using anti-Mbd3 antibody, but not control IgG, co-precipitated Brg1 (**Figure 2-2E**), indicating that these proteins interact in ES cells, as observed in K562 cells (Mahajan et al., 2005) and suggested by mass spectrometry of esBAF (Ho et al., 2009b). To further probe the functional relationship between Brg1 and Mbd3, we mapped Mbd3 genome-wide in Brg1 KD ES cells (**Figure 2-2F**), finding that Mbd3 localization was lost upon Brg1 KD. Together, these results indicate that Brg1 and Mbd3 directly associate with and regulate a common set of target genes in ES cells.

Mbd3 binding is enriched at bivalent genes

We next compared Mbd3 binding to previously published ChIP-Seq datasets in murine ES cells. **Figure 2-3A** shows the difference in Mbd3 binding levels between genes bound by a particular factor and those unbound, as defined in (Kim et al., 2010a). Mbd3 binding was substantially higher at targets of the Polycomb proteins Ezh1, Suz12, Phc1, and Eed, as well as the histone modifications associated with most Polycomb targets in ES cells: H3K27me3 and H3K4me3 (**Figures 2-3A-C**). Furthermore, similar to Polycomb localization, Mbd3 localization was strongest at CpG-rich promoters and weakest at CpG-poor promoters (**Figure 2-3D**). Finally, Mbd3 was localized to all four Hox

clusters, as previously observed for several PRC1 and PRC2 subunits (Boyer et al., 2006) (**Figure S2-3A**).

These data suggest that Mbd3 may function with Polycomb and/or MLL/SET1 complexes to regulate gene expression. Interestingly, we noted that the gene expression profile of Mbd3 KD ES cells strongly correlated with that of Suz12 KD ES cells, and to a lesser extent with that of Ash2l KD ES cells (**Figure 2-1A**), indicating that both H3K4me3 and H3K27me3 affect many "bivalent" genes similarly, and that Mbd3 may also regulate these genes. Indeed, most genes significantly up- or down-regulated upon Mbd3 KD were changed in the same direction upon KD of either Suz12 or Ash2l (**Figures 2-3E,F**), although the effects of Suz12 or Ash2l on gene expression were generally less severe. Thus, Mbd3 is directly associated with a significant fraction of Polycomb-bound genes, and contributes to their repression.

To attempt to understand why only a subset of Mbd3-bound genes were affected transcriptionally by Mbd3 knockdown, we carried out a multivariate regression analysis (see Experimental Procedures). Linear combinations of genome-wide binding data for 38 factors or modifications were optimized to best explain the effects of Mbd3 KD on expression of Mbd3/Brg1 common targets (**Figures S2-5A,B**). The full model exhibited a high correlation (R = 0.40) between predicted and measured effects of Mbd3 KD on gene expression. The strongest predictors of Mbd3 effects on gene expression included Tet1 and 5-hydroxymethylcytosine localization (see below). Beyond these, strong predictors of Mbd3 function included H3K27me3 (individual R=0.29), Mbd3 (R=0.13), and

transcription factors such as Esrrb, E2F1, and Klf4, whereas core pluripotency factors such as Oct4, Sox2, and Nanog contributed little (**Figure S2-5B** and data not shown). These data are consistent with proteomics studies showing that Esrrb physically associates with the BAF and Mbd3/NURD complexes (van den Berg et al., 2010), and suggest that genes regulated by this transcription factor are particularly sensitive to Mbd3 function.

Opposing chromatin remodeling functions of Mbd3 and Brg1 regulate recruitment of RNA Polymerase II

How do Mbd3 and Brg1 control gene expression at their common target genes? NURD complexes contain both a Swi2/Snf2-related ATPase (Mi-2) and a pair of histone deacetylase subunits (Hdac1 and Hdac2). We therefore analyzed nucleosome occupancy (measured by H3 abundance) and H4 acetylation at Mbd3/Brg1 targets by ChIP-qPCR. Genes repressed by Mbd3 exhibited decreased H3 occupancy and increased H4 acetylation upon Mbd3 KD (**Figures 2-4A,B**), suggesting that the Mbd3 complex represses target genes by deacetylating and stabilizing nucleosomes at promoters. Conversely, Brg1 knockdown resulted in variable effects on H4 acetylation, but consistent increases in H3 occupancy (**Figures 2-4A,B**). Thus, we find that Brg1 and Mbd3 antagonistically control nucleosome occupancy at target genes, with Brg1-mediated nucleosome loss associated with gene activation, and competing nucleosome stabilization and deacetylation by Mbd3 associated with gene repression.

How do these changes in chromatin architecture affect transcription? Many genes in eukaryotes are regulated by transcriptional pausing (Guertin et al., 2010). Given the localization of Mbd3 downstream of the TSS (**Figure 2-2C**), we considered the hypothesis that Mbd3 could play a role in enforcing a transcriptional pause. We therefore carried out whole-genome RNA Polymerase II (Pol II) mapping in EGFP KD, Mbd3 KD, and Brg1 KD ES cells. Overall, genome-wide Pol II localization was similar in all three KDs (**Figure 2-4C**), exhibiting in each case the expected promoter-proximal peaks corresponding to bidirectional transcription (Core et al., 2008; Seila et al., 2008). We next focused on the group of genes repressed by Mbd3 and activated by Brg1 (**Figure 2-4D**). If Mbd3 functioned solely to arrest Pol II, then mRNA increases in Mbd3 KD should be reflected in a loss of 5' Pol II and an increase in downstream Pol II. Instead, we find that the 5' peak of Pol II increases at these target genes in the Mbd3 KD, and decreases in the Brg1 KD, mirroring the effects of each KD on mRNA levels. Since the global Pol II profiles are nearly identical for all three KDs (**Figure 2-4C**), these data are consistent with a primary role for Mbd3 and Brg1 in regulating recruitment of Pol II specifically to target promoters, although an additional role in regulating Pol II pausing cannot be ruled out.

Knockdown of Tet1 impairs Mbd3 recruitment to target genes

Despite being named Methyl Binding Domain based on its homology to the methyl binding domain from MeCP2, mammalian Mbd3 does not appear to associate with cytosine-methylated DNA (Hendrich and Bird, 1998; Zhang et al., 1999). Interestingly, we noted that one prominent phenotype of Mbd3 null or KD

ES cells — upregulation of genes involved in trophectoderm differentiation — shares similarities to that of ES cells depleted of Tet1 (Ito et al., 2010; Koh et al., 2011), which converts 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC). We therefore speculated that Mbd3 might be regulated by 5hmC, rather than 5mC. To test this hypothesis, we compared Mbd3 localization to maps of Tet1 in murine ES cells (Wu et al., 2011c). Similar to Tet1 localization, Mbd3 was found at CpG-rich promoters, and correlated with Polycomb localization patterns (**Figures 2-3C-D**). High levels of 5hmC accompany Mbd3-associated genes in ES cells (Wu et al., 2011a) (**Figure 2-5A**), and Mbd3 and Tet1 localization patterns exhibited strong overlap (**Figure 2-5B**). Furthermore, genes repressed by Mbd3 were associated with significantly higher levels of Tet1 and 5hmC than genes unaffected by Mbd3 KD (**Figures S2-5B-C**).

We therefore asked whether cytosine hydroxymethylation functions in Mbd3 localization. Since, in ES cells, Tet1 KD significantly decreases hydroxymethylation levels, we carried out genome-wide mapping of Mbd3 in Tet1 KD ES cells. Strikingly, Mbd3 was completely delocalized in Tet1 KD cells, despite normal levels of Mbd3 protein in this knockdown (**Figures 2-5B-E**). Thus, hydroxymethylation, or some other aspect of Tet1 function, is necessary for Mbd3 association with target genes.

Mbd3 binds 5hmC- but not 5mC-containing DNA *in vitro*

Our results suggest that hydroxymethylation is not solely an intermediate state in a cytosine demethylation pathway, but that it also plays a distinct role in gene regulation via Mbd3 recruitment. As Mbd3 does not specifically bind to 5mC

in vitro, we considered the hypothesis that it might instead bind to 5hmC. To test this, we purified Mbd3/NURD complex from ES cells (**Figures 2-6A, S2-6A-B**), finding that this complex contained low levels of Tet1 (**Figure 2-6B**), identifying a direct physical link between Tet1 and Mbd3.

We carried out electrophoretic mobility shift assays (EMSAs) using probes with unmodified cytosine (C), 5mC, or 5hmC. As shown in **Figure 2-6C**, Mbd3/NURD complex had little effect on the methylated probe, but strongly shifted the 5hmC probe, whereas an untagged purification did not shift any of the probes (**Figure S2-6C**). Interestingly, Mbd3/NURD also shifted the unmodified "C" probe, but the shifted band was broader than the uniform shift of the 5hmC probe. These data suggest that Mbd3/NURD complex employs distinct modes of binding to unmodified and hydroxymethylated DNA.

To determine whether Mbd3/NURD's specificity for 5hmC over 5mC was due to Mbd3, we tested recombinant Mbd3 for binding to the same probes. Consistent with the results obtained using the full complex, recombinant Mbd3 exhibited strongest binding to hydroxymethylated DNA probes, with comparable but slightly reduced binding to unmodified probe, and dramatically lower binding to methylated probe (**Figures 2-6D, S2-6D**). As a control, recombinant methyl binding domain from Mbd1 specifically shifted the methylated probe (**Figure 2-6E**), as expected (Hendrich and Bird, 1998; Ohki et al., 2001). Together, these data confirm the hypothesis that Mbd3 and Mbd3 complexes do not bind to DNA harboring 5mC, but can bind 5hmC-marked DNA in a manner qualitatively distinct from that of DNA containing unmodified cytosine.

Knockdown of Mbd3 or Brg1 affects global 5hmC levels

Given the physical interaction between Mbd3 and Tet1, we next asked if Mbd3 affected 5hmC *in vivo*. Compared to control cells, Tet1 KD ES cells exhibited reduced 5hmC levels by dot blotting and thin layer chromatography (**Figures 2-7A,C**), as previously reported (Koh et al., 2011; Wu et al., 2011a) (Ficz et al., 2011a). To our surprise, both Mbd3 KD and Brg1 KD ES cells also exhibited strong reductions in global 5hmC levels as assayed by dot blotting (**Figures 2-7A, S2-7**). This effect of Mbd3 KD on bulk 5hmC levels was independently quantified by thin layer chromatography (**Figures 2-7B,C**). To extend these results to *in vivo* targets of Mbd3, we measured the enrichment of 5hmC (Song et al., 2011) at several Mbd3 target genes. We observed similar losses of 5hmC at six loci in Tet1 KD and in Mbd3 KD cells (**Figure 2-7D**). Finally, we compared 5hmC staining of individual cells by immunofluorescence in control cells to cells depleted of Tet1, Mbd3, or Brg1. Consistent with the above results, we observed noticeably reduced 5hmC levels in most Tet1, Mbd3 and Brg1 KD cells compared to control **(Figure 2-7E)**. Therefore, we conclude that Mbd3 and Brg1 function globally to establish or maintain normal levels of 5hmC in ES cells.

DISCUSSION

The epigenetic control of stem cell pluripotency and self-renewal has been subject to intense scrutiny in recent years. Extensive attention has been given to

specific coactivators/corepressors such as Polycomb Repressive Complexes, whereas other important chromatin regulators have received less attention. Here, we focused on Brg1 and Mbd3, components of coactivator and corepressor complexes, the mechanisms by which they antagonistically regulate a group of common target genes, and their role in the biology of 5hmC.

Antagonistic control of gene expression by BAF and Mbd3/NURD complexes

Several hundred genes were antagonistically regulated by Brg1 and Mbd3, showing increased expression in Mbd3KD, decreased expression in Brg1KD, and wild type-like expression in double KDs. These results are of great interest in the context of the “bivalent” chromatin architecture seen in ES cells, in which the activation-associated histone mark H3K4me3 and the repression-associated H3K27me3 mark co-occur at a large number of “master regulator” genes involved in cell fate decisions (Bernstein et al., 2006). However, reduction of H3K4me3 in ES cells by KD of Ash2l (Fazzio et al., 2008b) or Dpy-30 (Jiang et al., 2011) does not cause self-renewal defects, nor does H3K4me3 appear to be important for expression of most genes in ES cells (Jiang et al., 2011) (**Table S2-1**). Indeed, our results with Ash2l and Suz12 knockdowns show that reduction of either H3K4me3 or H3K27me3 results, “paradoxically”, in similar effects on gene expression (**Figures 2-3E,F**). Conversely, we show here that Mbd3 and Brg1 exhibit opposing functional effects on RNA Polymerase II recruitment and gene expression at several hundred genes. Together, these data suggest that the juxtaposition of opposing histone modifications or chromatin regulators near the

regulatory sequences of shared target genes is a common regulatory strategy in ES cells.

Mbd3/NURD plays a central role in 5-hydroxymethylcytosine biology

The inability of Mbd3 to bind specifically to 5mC (Hendrich and Bird, 1998; Zhang et al., 1999) raises the question of what function the “methyl binding domain” of this protein serves. We noted that ES cells depleted of Tet1, which catalyzes the hydroxylation of 5mC to form 5hmC, exhibit phenotypes similar to those of Mbd3 KD ES cells (Kaji et al., 2006; Koh et al., 2011; Zhu et al., 2009). Intriguingly, one of the few sequence differences in the methyl binding domain between human/mouse Mbd3 and remaining Mbd family members is a substitution of a phenylalanine for a tyrosine (numbered Y34/F34 in human Mbd1/Mbd3). This sequence change is largely responsible for Mbd3’s lack of 5mC binding *in vitro* (Saito and Ishikawa, 2002). In the structure of Mbd1 bound to methylated DNA, Y34 is located immediately adjacent to the 5-methyl group of 5mC (Ohki et al., 2001) (**Figure S2-6E**). Thus, loss of the hydroxyl group (Y34F) at this position in Mbd3 could almost perfectly compensate for the additional hydroxyl group in 5hmC relative to 5mC, thus allowing direct binding of Mbd3 to this modification in a manner structurally analogous to the binding of remaining Mbd members to 5mC.

Consistent with this hypothesis, we found that Mbd3 localization patterns are similar to Tet1 localization patterns, and Mbd3 is enriched at genes with high levels of hydroxymethylation (Wu et al., 2011c). Like Tet1, Mbd3 is associated largely with CpG-rich promoters bound by Polycomb, and is required for normal

expression of many of these targets. Mbd3 localization requires Tet1, suggesting that hydroxymethylation plays a role in Mbd3 recruitment *in vivo*. Finally, we found that Mbd3 preferentially binds to 5hmC-containing probes relative to 5mC-containing probes. Thus, our data are most consistent with a model in which Tet1-catalyzed hydroxymethylation serves to recruit Mbd3/NURD complex, and thus Mbd3/NURD may be an effector that mediates some of the effects of hydroxymethylation on gene expression.

Interestingly, we found that purified Mbd3/NURD also bound to unmodified probe *in vitro*, but that this binding was qualitatively distinct from binding to the 5hmC probe: the shifted band was more discrete and of higher mobility when the probe was hydroxymethylated. One potential explanation for this could be the binding location of the complex on the probe, as probe bound at either end should have higher mobility than internally bound probe. At present it is unclear to us why Tet1KD, which results in incomplete loss of 5hmC *in vivo*, almost completely abolishes Mbd3 localization. Given the increase in 5mC observed in Tet1KD cells at loci that lose 5hmC (Ficz et al., 2011a; Wu et al., 2011a), and our observation that 5mC inhibits Mbd3 binding, we speculate that genomic regions with 5mC intermixed with 5hmC might be unfavorable for Mbd3 binding. Future studies will dissect the details of how Mbd3/NURD differentially interacts with unmodified and hydroxymethylated DNA, and with hemi-modified or symmetrically modified DNA.

Finally, we found that KD of either Mbd3 or Brg1 results in reduction of bulk levels of 5hmC *in vivo*. This could occur via several possible mechanisms:

as two examples, Mbd3 could bind to a region of hydroxymethylated DNA and recruit Tet enzymes to hydroxylate adjacent methylcytosines, or Mbd3 could bind to hydroxymethylated loci and protect them from further steps in a demethylation pathway. Either way, the extensive interdependency between these factors – Tet1 is necessary for Mbd3 localization, Mbd3 is necessary for cytosine hydroxymethylation – is reminiscent of other co-dependencies in chromatin pathways. For instance, Polycomb is required for H2A.Z incorporation at promoters of developmental genes in ES cells, and H2A.Z is in turn required for Polycomb binding (Creyghton et al., 2008). Dissecting the detailed mechanistic basis for such interdependencies is a challenging goal for studies on chromatin regulation.

Together, these data describe the first downstream effector that regulates expression of 5hmC-marked genes, suggesting that 5hmC plays a role in gene regulation beyond serving simply as an intermediate in a demethylation pathway. It will be of great interest in future studies to determine whether Mbd3 also plays a role in 5hmC biology in other contexts such as early embryos (Inoue and Zhang, 2011a; Iqbal et al., 2011), imprinting (Reese et al., 2007), or neurons (Kriaucionis and Heintz, 2009). Finally, our discovery of an interdependent regulatory network consisting of 5hmC and two antagonistic chromatin regulators suggests that control of gene expression by opposing chromatin regulators is a common regulatory strategy in pluripotent ES cells.

Materials and Methods

RNAi-Mediated Knockdown

RNA interference using endoribonuclease III digested siRNAs (esiRNAs) was performed as described previously (Fazzio et al., 2008b), using E14 mouse ES cells. All KDs were performed for 48 hours to achieve effective KD, but avoid some indirect effects of prolonged loss of chromatin regulators. Stable Mbd3 KD lines were made by infection of ES cells with lentiviral shRNA vectors from the TRC library (Open Biosystems).

Expression Profiling

Array hybridizations were performed at the Sandler Asthma Basic Research (SABRE) Center Functional Genomics Core Facility as described previously (Fazzio et al., 2008a). For each single and double KD, a linear model was fit to the comparison to estimate the mean log2 (fold change) in 2 biological replicates and to calculate a moderated t-statistic, B statistic, false discovery rate and p-value for each probe. Adjusted p-values were produced as decribed previously (Holm, 1979). All procedures were carried out using functions in the R package limma (Gentleman et al., 2004; Smyth, 2004) or made4 (Culhane et al., 2005) in R/Bioconductor. Enrichment of Gene Ontology terms was performed with DAVID 6.7 (Dennis et al., 2003). mRNA data are available at GEO, accession GSE31008.

ChIP-Seq

Chromatin immunoprecipitation and deep sequencing library construction were performed using minor modifications of established protocols (Barski et al., 2009; Lee et al., 2006b) (Supplementary Experimental Procedures). Data are available at GEO, accession GSE31690.

Multivariate linear regression model

Multivariate linear regression was used to predict factors that regulate Mbd3/Brg1-target genes. 38 ChIP-seq datasets were divided into 40 bp bins surrounding the TSS of each gene that changed oppositely upon Mbd3 KD and Brg1 KD, with another bin representing the gene body. Binding of each of the 38 factors was compared to gene expression changes (see Supplementary Experimental Procedures for details).

Mbd3/NURD Purification

An ES cell line with a C-terminal 6-Histidine-3X FLAG tag placed just upstream of the *Mbd3* stop codon was constructed as described in the Supplementary Experimental Procedures. Mbd3/NURD complex was purified from ~ 4 X 10^8 of these cells by sequential affinity purification steps using FLAG-M2 Agarose followed by TALON Agarose beads.

EMSA Assays

EMSA assays were performed on 5% polyacrylamide 0.5X TBE gels. Biotinylated probes corresponding to +529 - +628 of the *Tgfb1* gene were made by PCR using Phusion polymerase and NTPs containing either unmodified, methylated, or hydroxymethylated dCTP. The PCR primer sequences for probe construction were: biotin-TGCCTCTTGAGTCCCTCGCATC and AGTGGGTGTTCTTAAATAGGGGAGCT. Binding reactions were performed in 1X Binding Buffer (LightShift Chemiluminescent EMSA Kit; Thermo Scientific) with 100 mM KCl, 8% glycerol, 0.02% NP-40, 5 mM $MgCl_2$, 0.85 μg BSA, 30 ng yeast genomic DNA, 1 mM ATP, and with or without 1 μl purified Mbd3/NURD or recombinant Mbd3, as indicated. After electrophoresis, samples were transferred to charged Nylon membrane and probed with streptavidin-HRP according to the instructions in the EMSA kit.

Dot Blotting

2-fold serial dilutions of genomic DNA were denatured in 0.4 M NaOH/10mM EDTA at 95C for 10 min, then added to an equal volume of cold 2M ammonium acetate (pH 7.0). Denatured DNA samples were spotted onto nitrocellulose. The membrane was washed with 2XSSC buffer and then UV cross-linked. Membrane was blocked with 5% non-fat milk for 1 hr and incubated with rabbit anti-5hmC, detected by HRP-conjugated secondary antibody and enhanced chemiluminescence.

Immunofluorescence

Cells were resuspended in PBS, cytospun onto glass slides, fixed in 4% paraformaldehyde for 10 min at room temperature, washed twice with PBS, permeabilized in 2% TritonX-100-PBS for 10 min, and blocked with 5% FBS and 0.1% TritonX-100 in PBS for 15 min at room temperature. Cells were incubated with primary antibody diluted 1:500 with 5% FBS and 0.1% Triton X-100 in PBS for 30 minutes at room temperature, washed 3 X 5 min with PBS, and incubated for 30 min at room temperature with secondary antibody diluted 1:200 with 5% FBS and 0.1% Triton X-100 in PBS. DAPI was added in mounting medium (Vectashield, H1500).

Supplemental Experimental Procedures

Chromatin Immunoprecipitation

80% Confluent ES cells were crosslinked by addition of formaldehyde (Sigma) to a final concentration of 1% and incubated at room temperature (RT) for 10 min. Crosslinking was quenched with 125mM Glycine. Cells were washed twice with ice cold PBS containing PMSF, collected in PBS, and pelleted at 1000g for 5 min, at 4^{o}C. Cell pellets were either flash frozen in liquid nitrogen and stored in -80^{o}C, or processed immediately. Pellets were resuspended in 270 µl SDS-Lysis Buffer (1% SDS, 10mM EDTA and 50mM Tris-Cl, pH 8.1.) including protease inhibitor complex (Sigma) and PMSF (Sigma). Samples were sonicated in a Bioruptor (UCD-200) at high setting 3 times for 15 min of 30sec on/30sec off cycles, followed by a 14000 rpm spin at 4^{o}C for 1 hour. Supernatants were

transferred to a new tube and pellets were resuspended in 150 µl of lysis buffer and processed in a Bioruptor using the previous settings for two more 15 min cycles. Supernatants were collected after a 13000 rpm spin at 4^{o}C for 10 min and combined with supernatants from the previous centrifugation. Chromatin was quantified by measuring A_{260}. 50 µg of protein A magnetic beads (NEB) were washed twice with 600 µl of 0.1M sodium phosphate (NaK) buffer (pH 8.0) and antibody was coupled in 200 µl NaK buffer for 30min at RT. Antibody coupled beads were blocked with 5mg/ml BSA-PBS solution for 1hr at 4^{o}C. 70 µg of chromatin for each immunoprecipitaton was diluted 1:10 in IP-Blocking buffer (5mg/ml BSA and 5µg/ml yeast tRNA in 0.01% SDS, 1.1% Triton X-100, 1.2mM EDTA, 16.7mM Tris-Cl (pH 8.1), 167mM NaCl), combined with antibody coupled magnetic beads and incubated at 4^{o}C overnight (up to 16 hours). Magnetic beads were washed once with RIPA buffer (0.1%SDS, 10mM Tris (pH 7.6), 1mM EDTA, 0.1% Na-Deoxycholate, 1% Triton X-100); 3 times with NaCl buffer (0.5M NaCl in RIPA); 3 times with LiCl buffer (0.3 M LiCl, 0.5% NP40, 0.5% Na-Deoxycholate in RIPA); twice with 0.2% TritonX-100 TE buffer and once with TE buffer (pH 8.0).

Washed beads were resuspended in 300 µl TE and incubated at 65^{o}C overnight in after adding the following (final concentrations): SDS (3%), proteinaseK (1mg/ml) (Sigma) and glycogen (0.5mg/ml) (Ambion). Eluted material was transferred to a new tube and the beads were resuspended in 150 µl 0.5M NaCl-TE solution and incubated at 65^{o}C for one hour. Eluted materials were combined and PCI (Phenol-Chloroform-Isoamylalcohol) extracted using

phase-lock tubes (Eppendorf). Ethanol precipitated ChIP DNA was treated with RNAse (Qiagen) for 2 hours and with CIP for an hour at 37^{o}C (NEB calf alkaline phosphatase - 0.25U/µl in 1x NEB Buffer 3). Reactions were then cleaned up with Qiagen MiniElute spin columns.

Antibodies

The following antibodies were used in this study for ChIP, IP, Western blotting, dot blotting and meDIP: Mbd3, Abcam (ab3755) and Bethyl (A302-528A); Brg1, Bethyl (A300-813A); Pol II, Santa Cruz (sc-899X); Ac-H4, Millipore (06-866); H3, Abcam (ab1791); Tet1, Millipore (09-872); 5hmC, Active Motif (39791); 5mC, Eurogentec (BI-MECY-0500); Mta1, Bethyl (A300-280A); and beta-actin, Sigma (A1978).

Generation of stable Mbd3 knockdown lines

80 - 90 % confluent 293T cells were transfected with 15µg of shRNA containing vector, 20µg of pCMVΔ8.9 and 3µg of pVSV-G using Lipofectamine2000 in OPTI-MEM. The viral supernatant was collected at 48h and 72h and filtered with low protein binding 0.2µM syringe filters. To enrich the virus yield, virus producing cells were lysed with 2 cycles of freeze-thaws using liquid nitrogen and a 37^{o}C water bath. After pelleting the cellular debris with centrifugation, the supernatant was filtered and combined with the filtered growth medium. Virus was concentrated with ultracentrifugation at 4 ^{o}C, 27 000 rpm for 2 hr in an SW28 rotor. The viral pellet was resuspended in 15 ml HBS and ultra filtered with Ultracell-100k filters (Amicon). ES cells were infected in 2ml ES

medium containing 6mg/ml Polybrene® in a 6well-plate with 100ul of the concentrated virus. The next day, cells were re-plated on 10cm dishes and antibiotic selection started 24hr after the infection. Resistant colonies were picked, expanded and tested for knockdown efficiency via RTqPCR and Western blotting.

Isolation of hydroxymethylated DNA

Hydroxymethylated DNA was isolated as previously described (Song et al., 2011). Briefly, purified genomic DNA, sonicated into short fragments (200-500 bp), was incubated with β-glucosyltransferase (β-GT) and UDP-6-N_3-Glucose, which resulted in N_3-glucosylated-5hmC. Modified DNA was then biotinylated using copper-free click chemistry. Biotinylated DNA was isolated using streptavidin magnetic beads and eluted in 100mM DTT for 2hrs at room temperature. Isolated DNA was cleaned up using Micro Bio-Spin 6 spin columns (Bio-Rad) followed by MinElute reaction clean up columns (Qiagen) and subjected to qPCR.

Deep sequencing library construction

ChIP material was pre-size selected (100 to 800bp in size) on a 2% TAE agarose gel using the MiniElute columns QiaQuick (Qiagen). Gel purified DNA fragments were blunt ended and phosphorylated with the EPICENTRE End-it-Repair kit (1X buffer, 0.25mM dNTPs,1mM ATP, 1ul/50ul reaction of Enzyme mix) for 1hr at RT and cleaned up with Qiagen MiniElute spin columns. Adenosine nucleotide overhangs were added using EPICENTRE exo- Klenow for

45min at RT (with 0.2mM dATP). Illumina genome sequencing adaptors were then ligated using the EPICENTRE Fast-Link ligation kit: 11.5µl A tailed DNA eluted from a MinElute column was mixed with 1.5µl 10X ligation buffer, 0.75µl 10mMATP, 0.5µl Illumina DNA adaptors and 1µl ligase. The reaction was incubated for 1hr at RT and subsequently supplemented with 7.5 µl water, 1µl 10X buffer, 0.5µl 10mM ATP and 1µl ligase, and incubated overnight at 16^{o}C.

The ligation reaction was cleaned up with MiniElute columns (with an additional wash step to eliminate all the excess adaptors) and the adaptor ligated fragments were amplified by PCR as follows:

0.75 µl of each Illumina genomic DNA sequencing primers, 6µl 10xPfx buffer 1.8µl 10mM dNTPs, 1.2µl 50mM MgSO4 and 1µl Pfx DNA polymerase (Invitrogen) were added to 30µl DNA template in a 100ul reaction. The cycling parameters were: 1. 94oC 2'; 2. 94oC 15''; 3. 65oC 1'; 4. 68oC 30''; 5. repeat from step 2 16 times; 6. 68oC 5'. The PCR product was size selected (250 to 450 bp) from a 2% TAE agarose gel using QiaQuick columns (Qiagen). Gel purified fragments were finally precipitated with Sodium acetate and ethanol and pellets were resuspended (25nM final concentration) in TE buffer and subjected to SOLEXA single-read sequencing at the UMass Medical School deep sequencing core facility.

Mbd3 datasets were collected for two independent ChIP samples using independent antibodies (obtained from Bethyl and Abcam). Datasets were well-correlated (**Figure S7**), although the Bethyl library exhibited higher genome-wide background than the Abcam library and so was not used for most analyses.

Nonetheless, all major biological conclusions were reproduced with either dataset.

Multivariate Adaptive Regression Splines (MARS) Model

MARS models were used to predict the log2 expression fold change upon Mbd3 knockdown using 38 ChIP-seq datasets (log2 of histone modification and transcription factor binding; called features). Each gene is divided into 41 bins, 40 100-bp bins for its promoter region ([-2k, +2k] around TSS) and one bin for the remaining gene body (from +2k to the transcriptional termination site), and the bin that leads to the highest correlation coefficient is chosen as representative signal for that gene. Feature selection was performed and the final model includes 12 features. A MARS model is judged by Person's correlation coefficient (R) between the predicted and experimental log2 expression change at 10-fold cross-validation. The MARS modeling was performed using the earth() function in R, and the relative importance of a histone mark or binding of a transcription factor in the final model is calculated using two criteria: percentage of models that use the feature during feature selection, and generalized cross validation (GCV) (Friedman, 1991). We constructed the model with the 568 transcripts on which knocking down Brg1 and Mbd3 had opposite effects (Brg1 upregulated and Mbd3 downregulated, or Brg1 downregulated and Mbd3 upregulated) that were significant in both knockdown datasets (adjusted p-value < 0.01).

Gene Targeting

A targeting construct containing one homology arm extending from ~2.5 kb upstream of the Mbd3 transcriptional termination site (TTS) to just downstream of the TTS and a second homology arm extending from just downstream of the TTS to ~3.2 kb further downstream was constructed. Sequence encoding 36 amino acids (SGRGSHHHHHHAGMDYKDHDGDYKDHDIDYKDDDDK) that include a 6-Histidine and a triple FLAG tag was inserted just upstream of the normal Mbd3 stop codon. A LoxP-Hygromycin-LoxP selection cassette was introduced at the upstream end of the second homology arm and a diptheria toxin A (DTA) counterselection cassette just downstream of the second homology arm (**Figure S2-6A**). 25 g of the resulting plasmid was linearized with Asc I and electroporated into E14 ES cells. Clones were screened by PCR, followed by western blotting to ensure correct integration (**Figure S6B**).

Mbd3/NURD Purification

Approximately $4X10^8$ Mbd3-6His-3XFLAG ES cells were collected after trypsinization from eight 15 cm dishes, washed once in PBS and lysed by sonication in Lysis buffer (50mM Tris, pH 7.4, 250mM NaCl, 0.1% Tx-100), plus 1X HALT protease inhibitors (Thermo). The lysate was clarified twice by centrifugation and incubated for three hours in 200 l of FLAG M2 Agarose bead slurry (Sigma) that had previously been equilibrated in Lysis buffer. Beads were washed five times and Mbd3/NURD complex was eluted five times with 0.5

mg/ml 3XFLAG peptide (Sigma) in Lysis Buffer, with 15 minute incubations prior to recovery of each elution. Pooled FLAG elutions were incubated with 100 l of pre-equilibrated TALON Agarose beads for two hours, washed five times with Lysis buffer, and eluted five times in Lysis buffer + 250 mM imidazole (10 minute incubations per elution). Pooled eluates from the TALON beads were dialyzed overnight in Buffer H (25 mM Hepes-KOH pH 7.6, 0.1 mM EDTA, 0.5 mM EGTA, 2 mM MgCl2, 20 % glycerol, 0.02 % NP40) with 100 mM KCl, aliquoted, flash-frozen in liquid nitrogen and stored at -80 ^{o}C.

Purification of recombinant Mbd3 or Mbd1-MBD

The longest isoform of Mbd3 (Mbd3a) and the first 75 amino acids (comprising the methylcytosine binding domain) of Mbd1 were PCR amplified from mouse ES cell cDNA with Phusion DNA polymerase, cloned into pCR2.1, sequence verified, sub-cloned into pQE80L in frame with an N-terminal six histidine tag encoded within the vector and transformed into BL21 pLysS cells. 200 ml cultures were grown to early log phase, induced with 0.5 mM IPTG for 3.5 hours, washed with PBS and frozen. Frozen pellets were re-suspended in lysis buffer (see Mbd3/NURD purification above), sonicated several times and centrifuged twice to clarify lysates. Lysates were incubated with 200 l of TALON resin for 2 hours, washed 6 times with 1 ml lysis buffer and eluted 5 times in lysis buffer containing 250 mM imidazole. Purified proteins were dialyzed overnight in Buffer H (see above), aliquoted and flash frozen in liquid nitrogen.

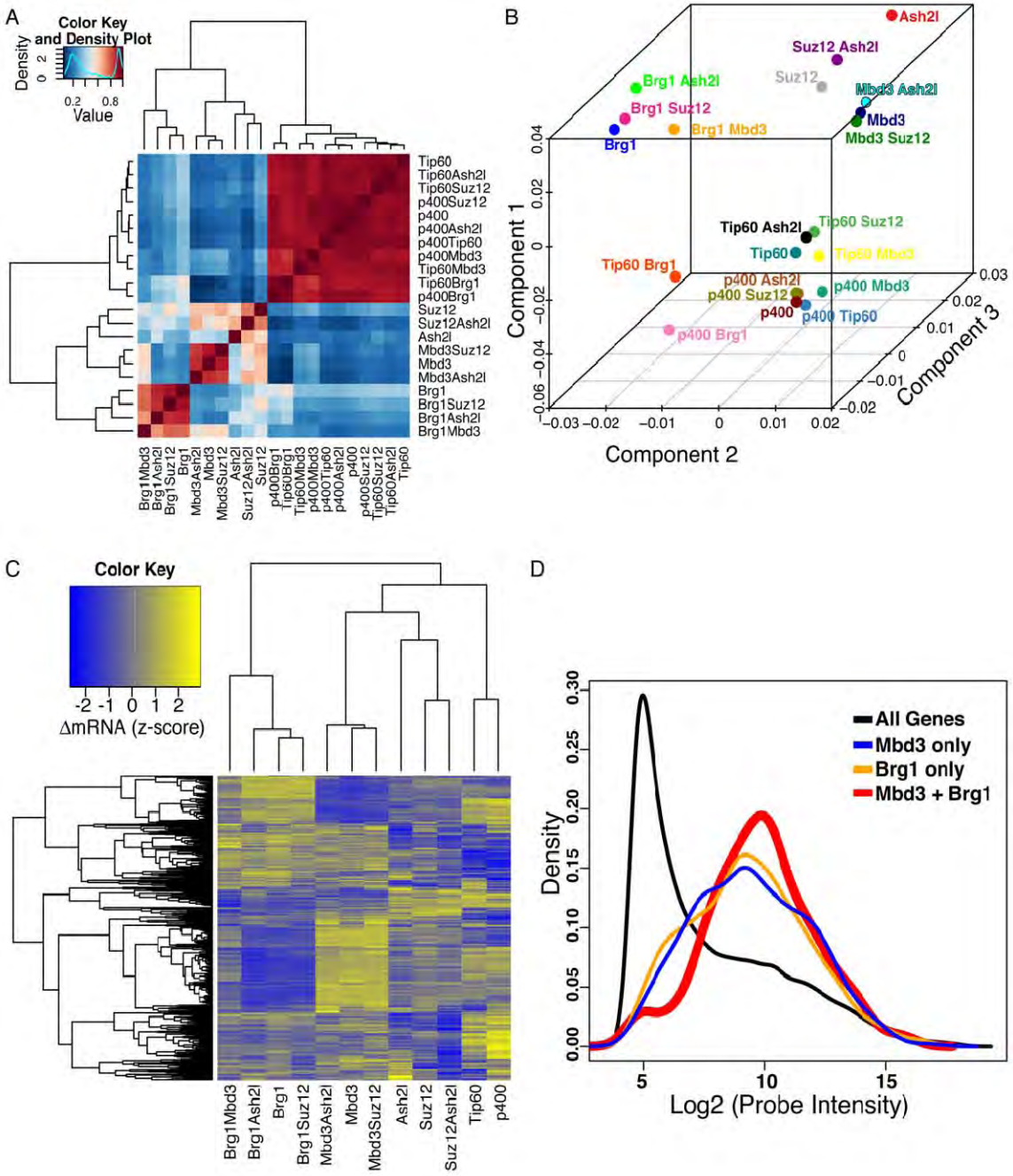

A
Color Key and Density Plot
Density
Value
Tip60
Tip60Ash2l
Tip60Suz12
p400Suz12
p400
p400Ash2l
p400Tip60
p400Mbd3
Tip60Mbd3
Tip60Brg1
p400Brg1
Suz12
Suz12Ash2l
Ash2l
Mbd3Suz12
Mbd3
Mbd3Ash2l
Brg1
Brg1Suz12
Brg1Ash2l
Brg1Mbd3
B
Ash2l
Suz12 Ash2l
Brg1 Ash2l
Suz12
Mbd3 Ash2l
Brg1 Suz12
Brg1 Mbd3
Mbd3
Brg1
Mbd3 Suz12
Tip60 Ash2l
Tip60 Suz12
Tip60
Tip60 Mbd3
Tip60 Brg1
p400 Ash2l
p400 Suz12
p400 Mbd3
p400
p400 Tip60
p400 Brg1
Component 1
Component 2
Component 3
C
Color Key
ΔmRNA (z-score)
Brg1Mbd3
Brg1Ash2l
Brg1
Brg1Suz12
Mbd3Ash2l
Mbd3
Mbd3Suz12
Ash2l
Suz12
Suz12Ash2l
Tip60
p400
D
All Genes
Mbd3 only
Brg1 only
Mbd3 + Brg1
Density
Log2 (Probe Intensity)

Figure 2-1

Figure2-1. Antagonistic effects of Brg1 and Mbd3 on gene expression in mES cells.

(A) Gene expression data for single and double knockdowns of Brg1, Mbd3, Ash2l, Suz12, p400, and Tip60. Heatmap shows pairwise Pearson correlation coefficients for the 21 datasets. Four major clusters emerge, roughly corresponding to Brg1, Mbd3, Ash2l/Suz12, and Tip60/p400.

(B) Principal Component Analysis. Genes significantly misregulated (adjusted p-value < 0.01) in any data set from (**A**) were subjected to principal component analysis. Shown are individual data sets plotted along three most prominent principal components, which account for 87% of the total variance in gene expression.

(C) Mbd3 and Brg1 antagonistically regulate a common set of genes. Unsupervised clustering of genes misregulated in both the Mbd3 KD and Brg1 KD datasets (adjusted p-value < 0.05). Clustering was performed on data sets containing either Mbd3 KD or Brg1 KD (or both), as well as the Tip60, p400, Suz12 and Ash2l single KD datasets for contrast.

(D) Genes regulated by both Mbd3 and Brg1 tend to be moderately to highly expressed. Shown are mRNA abundance distributions (expressed as log2 of microarray probe intensity) for all genes, and for genes regulated by either or both Mbd3 and Brg1.

See also **Figure S2-1**, **Table S2-1** and **Table S2-2**.

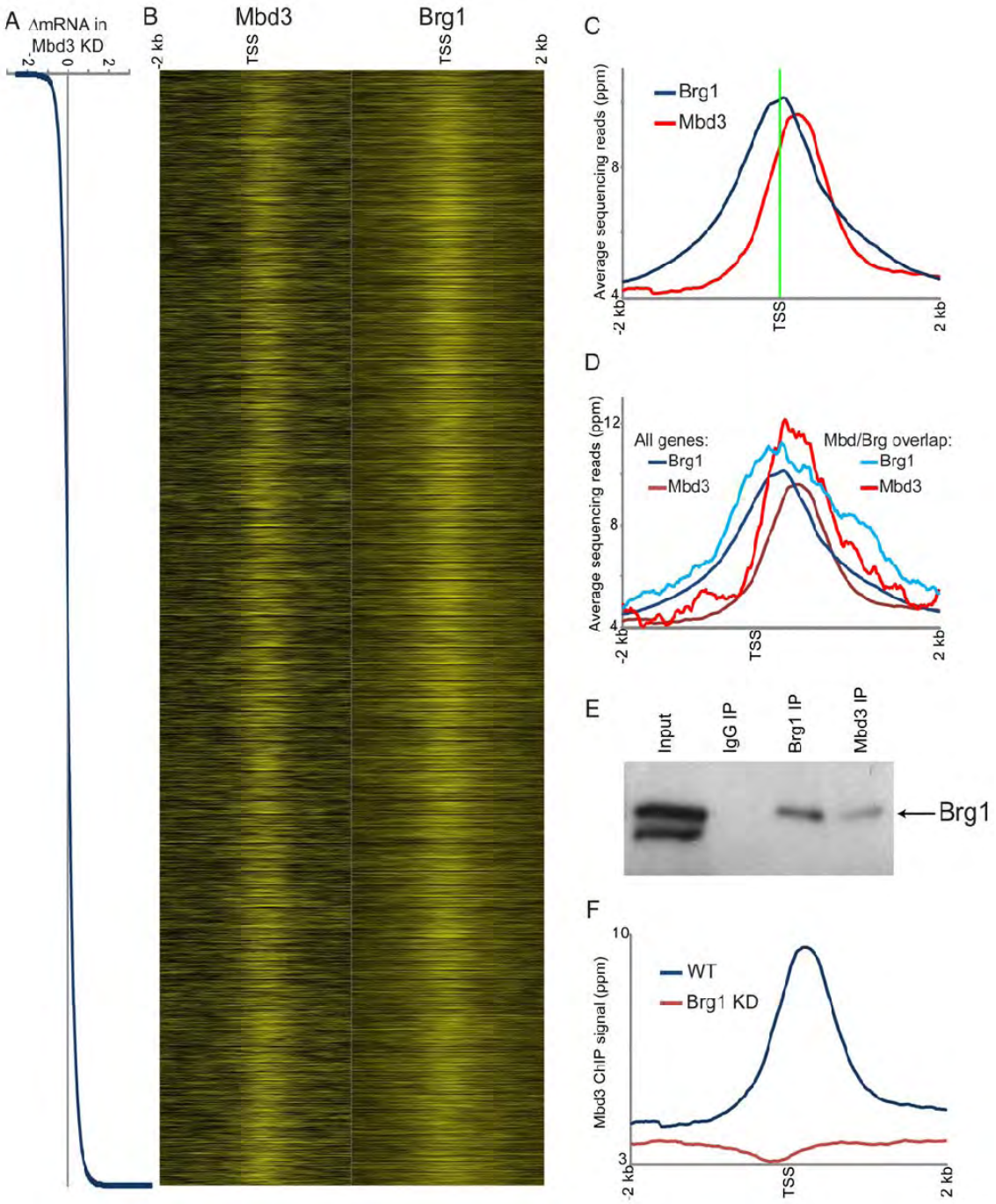

A
ΔmRNA in Mbd3 KD
-2
0
2
B
Mbd3
Brg1
-2 kb
TSS
TSS
2 kb
C
Brg1
Mbd3
Average sequencing reads (ppm)
8
4
-2 kb
TSS
2 kb
D
12
8
4
All genes:
Brg1
Mbd3
Mbd/Brg overlap:
Brg1
Mbd3
Average sequencing reads (ppm)
-2 kb
TSS
2 kb
E
Input
IgG IP
Brg1 IP
Mbd3 IP
Brg1
F
10
3
WT
Brg1 KD
Mbd3 ChIP signal (ppm)
-2 kb
TSS
2 kb

Figure 2-2

Figure2-2. Genome-wide localization of Mbd3.

(A) Mbd3 KD effects on gene expression. Genes are sorted by change in mRNA abundance in Mbd3 KD, shown here in Log_2.

(B) Mbd3 was mapped across the genome in ES cells by ChIP-Seq. Left panel: Mbd3 mapping data for 4 kb surrounding the transcriptional start sites (TSS) of 17,992 genes for which Mbd3KD gene expression data were available, with heatmap yellow saturating at 20 ppm normalized abundance. Right panel: published data for Brg1 (Ho et al., 2009a). In both panels, genes are sorted as in (**A**).

(C) Mbd3 binds downstream of Brg1. Averaged data for all genes in (**B**) are shown relative to the TSS.

(D) Mbd3 and Brg1 physically associate with antagonistically-regulated genes. Mbd3 and Brg1 data are shown for all genes as in **C**, or only for genes significantly repressed by Mbd3 and activated by Brg1.

(E) Mbd3 and Brg1 physically associate. Western blots for Brg1 following immunoprecipitation with the indicated antibodies.

(F) Brg1 is required for Mbd3 localization. Mbd3 was mapped genome-wide in Brg1 KD cells, and data for all genes are averaged and plotted as in (**C**).

See also **Figure S2-2**, **Figure S2-3**, and **Figure S2-4**.

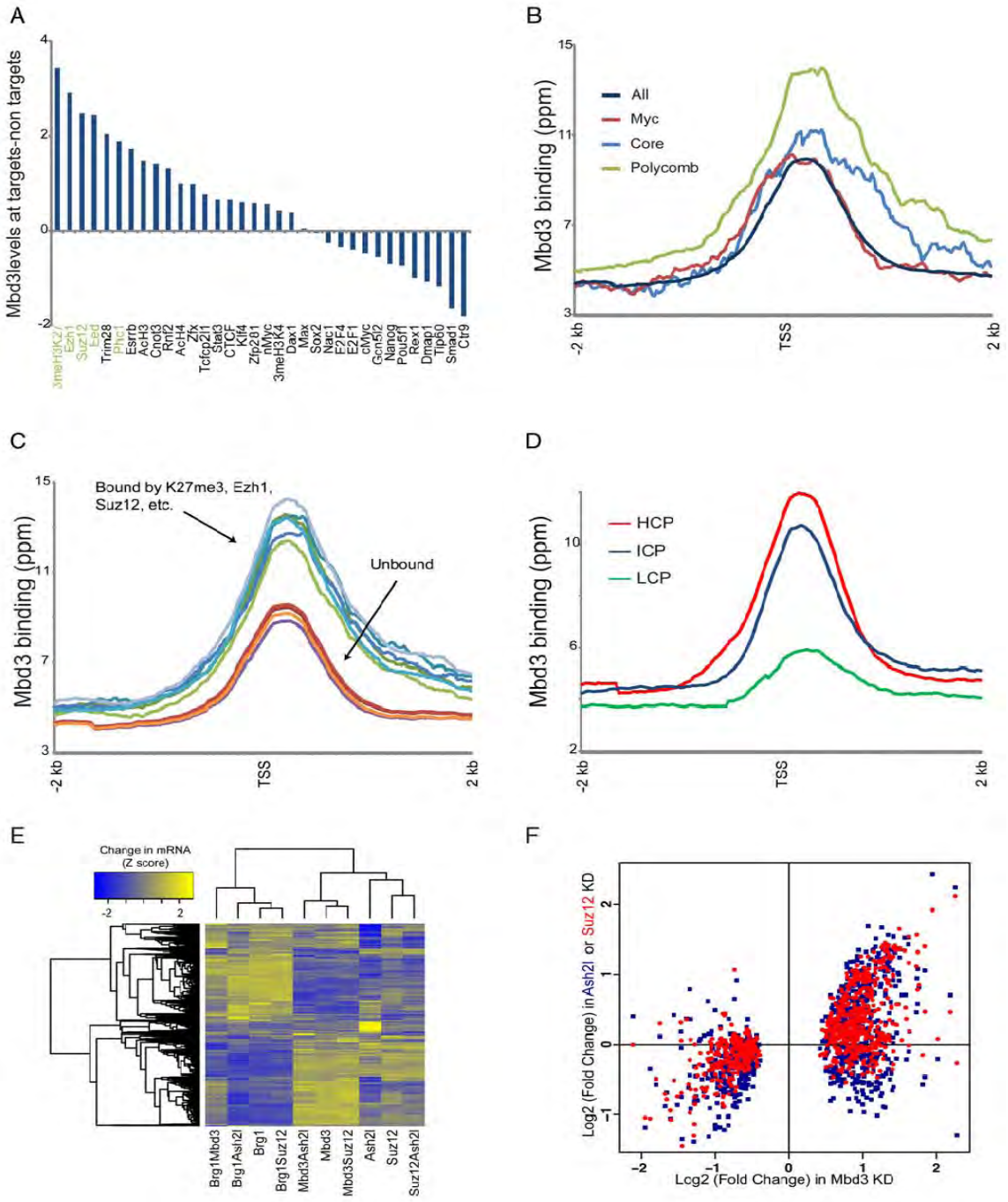

A
Mbd3levels at targets-non targets
4
2
0
-2
3meH3K27
Ezh1
Suz12
Eed
Phc1
Trim28
Esrrb
AcH3
Cnot3
Rnf2
AcH4
Zfx
Tcfcp2l1
Stat3
CTCF
Klf4
Zfp281
nMyc
3meH3K4
Dax1
Max
Sox2
Nac1
E2F4
E2F1
cMyc
Gcn5l2
Nanog
Pou5f1
Rex1
Dmap1
Tip60
Smad1
Ctr9
B
Mbd3 binding (ppm)
15
11
7
3
All
Myc
Core
Polycomb
-2 kb
TSS
2 kb
C
Mbd3 binding (ppm)
15
11
7
3
Bound by K27me3, Ezh1, Suz12, etc.
Unbound
-2 kb
TSS
2 kb
D
Mbd3 binding (ppm)
10
6
2
HCP
ICP
LCP
-2 kb
TSS
2 kb
E
Change in mRNA (Z score)
-2
2
Brg1Mbd3
Brg1Ash2l
Brg1
Brg1Suz12
Mbd3Ash2l
Mbd3
Mbd3Suz12
Ash2l
Suz12
Suz12Ash2l
F
Log2 (Fold Change) in Ash2l or Suz12 KD
2
1
0
−1
−2
−1
0
1
2
Lcg2 (Fold Change) in Mbd3 KD

Figure 2-3

Figure2-3. Mbd3 directly regulates Polycomb target genes.

(A) Mbd3 binding at Polycomb target genes. For 34 mapped factors with thresholded binding defined in (Kim et al., 2010a), Mbd3 binding levels (mean ChIP signal for 1 kb centered on the +200 position) were calculated for bound and unbound subsets of genes. Genes annotated as unbound by all 34 factors were removed from this analysis. Factors are sorted according to the relative Mbd3 binding at factor targets relative to nontargets.

(B) Mbd3 binding at three ES cell “modules.” Average Mbd3 binding for all genes, and for the three modules defined in (Kim et al., 2010a), is plotted relative to the TSS.

(C) Mbd3 binding at Polycomb targets and non-targets. For various Polycomb-related marks, averaged Mbd3 profile is shown for bound and unbound genes.

(D) Mbd3 binds preferentially to high-CpG promoters. Mbd3 binding data are averaged for high, intermediate, and low CpG (HCP, ICP, and LCP) promoters, as defined in (Weber et al., 2007).

(E) Mbd3 KD affects Polycomb targets. Clustered mRNA data for KD of Suz12, Ash2l, Mbd3, or Brg1, and assorted double knockdowns. Genes significantly misregulated in any of the included datasets are shown.

(F) Scatterplot of Mbd3 KD gene expression vs. Ash2l KD and Suz12 KD. Only genes showing significant misregulation in Mbd3 KD are shown.

See also **Figure S2-3** and **Figure S2-5**.

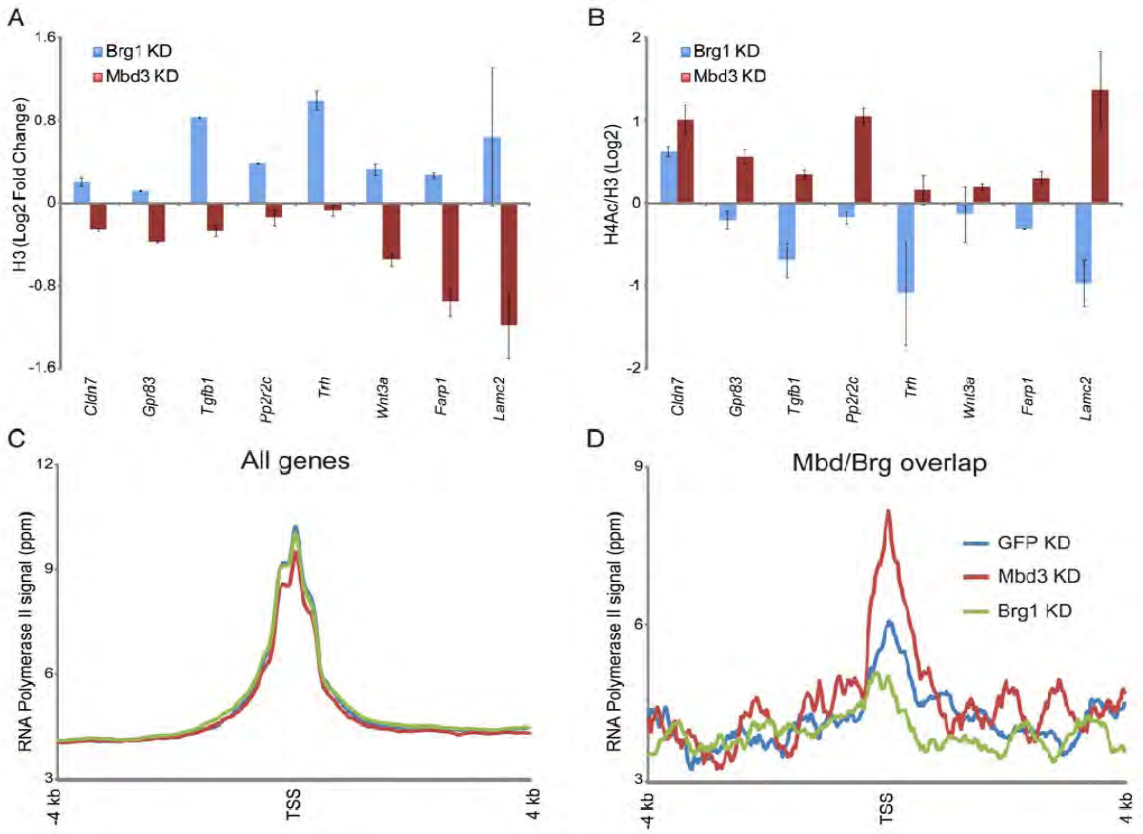

A
Brg1 KD
Mbd3 KD
H3 (Log2 Fold Change)
1.6
0.8
0
-0.8
-1.6
Cldn7
Gpr83
Tgfb1
Pp2r2c
Trh
Wnt3a
Farp1
Lamc2
B
Brg1 KD
Mbd3 KD
H4Ac/H3 (Log2)
2
1
0
-1
-2
Cldn7
Gpr83
Tgfb1
Pp2r2c
Trh
Wnt3a
Farp1
Lamc2
C
All genes
RNA Polymerase II signal (ppm)
12
9
6
3
-4 kb
TSS
4 kb
D
Mbd/Brg overlap
RNA Polymerase II signal (ppm)
9
6
3
GFP KD
Mbd3 KD
Brg1 KD
-4 kb
TSS
4 kb

Figure 2-4

Figure2-4. Mbd3 and Brg1 regulate chromatin structure and transcription initiation at target genes.

(A) H3 occupancy at Mbd3/Brg1 targets. H3 occupancy is plotted for control KD, Mbd3 KD, and Brg1 KD.

(B) H4 acetylation at Mbd3/Brg1 targets. As in (**A**), for H4 acetylation. Data are normalized to nucleosome occupancy (H3 ChIP).

(C) Genome-wide RNA Polymerase II (Pol II) mapping. Average TSS-aligned profiles of Pol II occupancy is shown for all genes for control, Brg1 KD, and Mbd3 KD cells.

(D) Pol II levels at the 5' end correlate with mRNA abundance changes in Brg1 and Mbd3 KDs. As in (**C**), but only for genes repressed by Mbd3 and activated by Brg1.

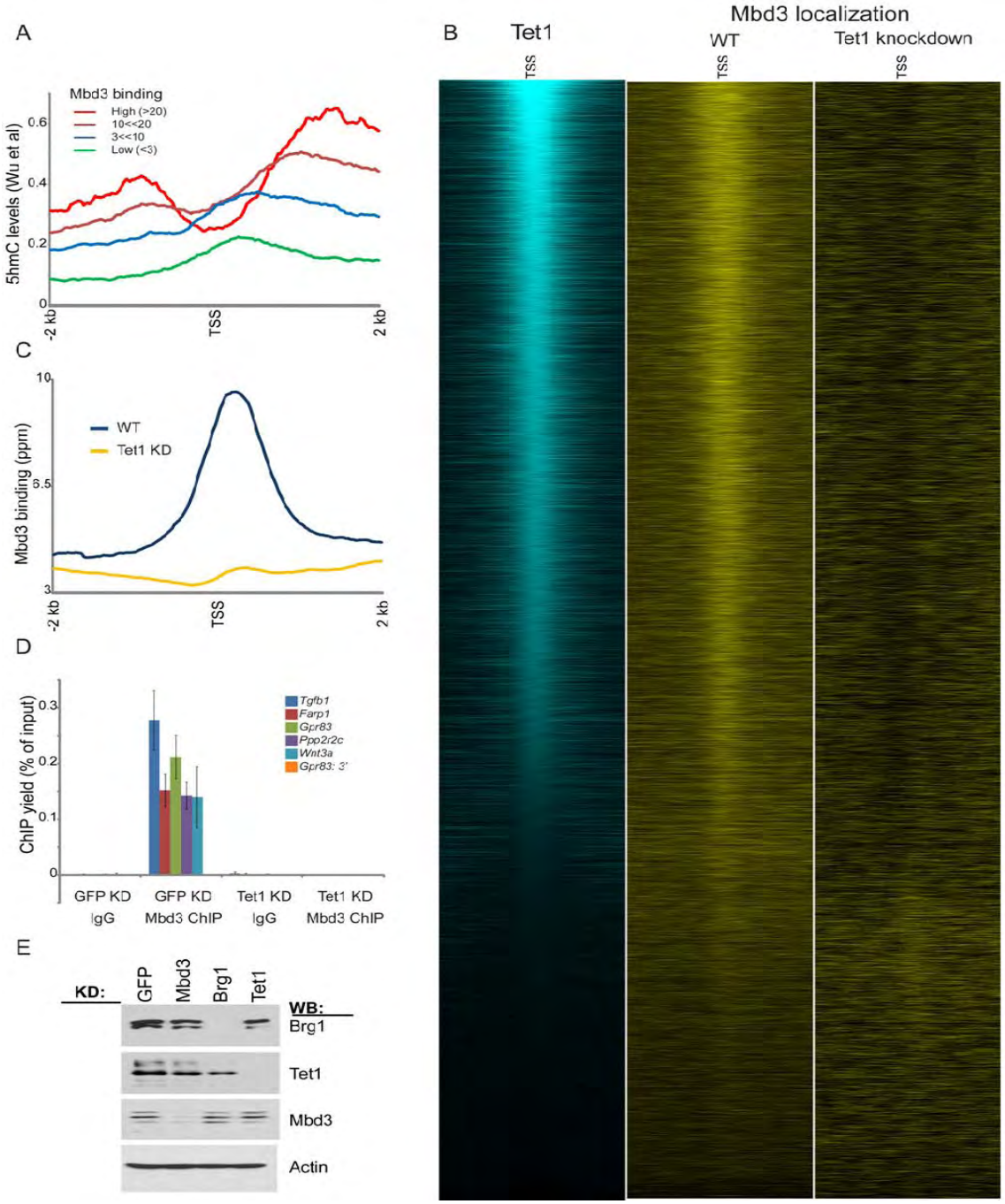

A
Mbd3 binding
High (>20)
10<<20
3<<10
Low (<3)
5hmC levels (Wu et al)
0.6
0.4
0.2
0
-2 kb
TSS
2 kb
B
Tet1
Mbd3 localization
WT
Tet1 knockdown
TSS
TSS
TSS
C
Mbd3 binding (ppm)
10
6.5
3
WT
Tet1 KD
-2 kb
TSS
2 kb
D
ChIP yield (% of input)
0.3
0.2
0.1
0
Tgfb1
Farp1
Gpr83
Ppp2r2c
Wnt3a
Gpr83: 3'
GFP KD IgG
GFP KD Mbd3 ChIP
Tet1 KD IgG
Tet1 KD Mbd3 ChIP
E
KD:
GFP
Mbd3
Brg1
Tet1
WB:
Brg1
Tet1
Mbd3
Actin

Figure 2-5

Figure2-5. Mbd3 associates with hydroxymethylated regions of the genome.

(A) Mbd3-bound genes are associated with high levels of hydroxymethylation. Average hydroxymethylation data from (Wu et al., 2011a) are averaged for genes with the indicated levels of Mbd3 binding.

(B) Mbd3 colocalizes with Tet1. Left panel: Tet1 (Wu et al., 2011c) mapping data are shown for all named genes, sorted by Tet1 binding levels. Middle and right panels: Mbd3 localization is shown for control and Tet1 KD ES cells.

(C) Mbd3 binding data for control and Tet1 KD are plotted as in **Figures 2-2C,D**.

(D) qPCR validation of Mbd3 binding at 6 selected target loci in GFP or Tet1 KD. Shown are mean +/- sem.

(E) Knockdowns of Tet1, Brg1, and Mbd3 do not affect protein levels of the other remaining factors. Knockdowns of the various factors were assayed by Western blot, as indicated.

See also **Figure S2-5**.

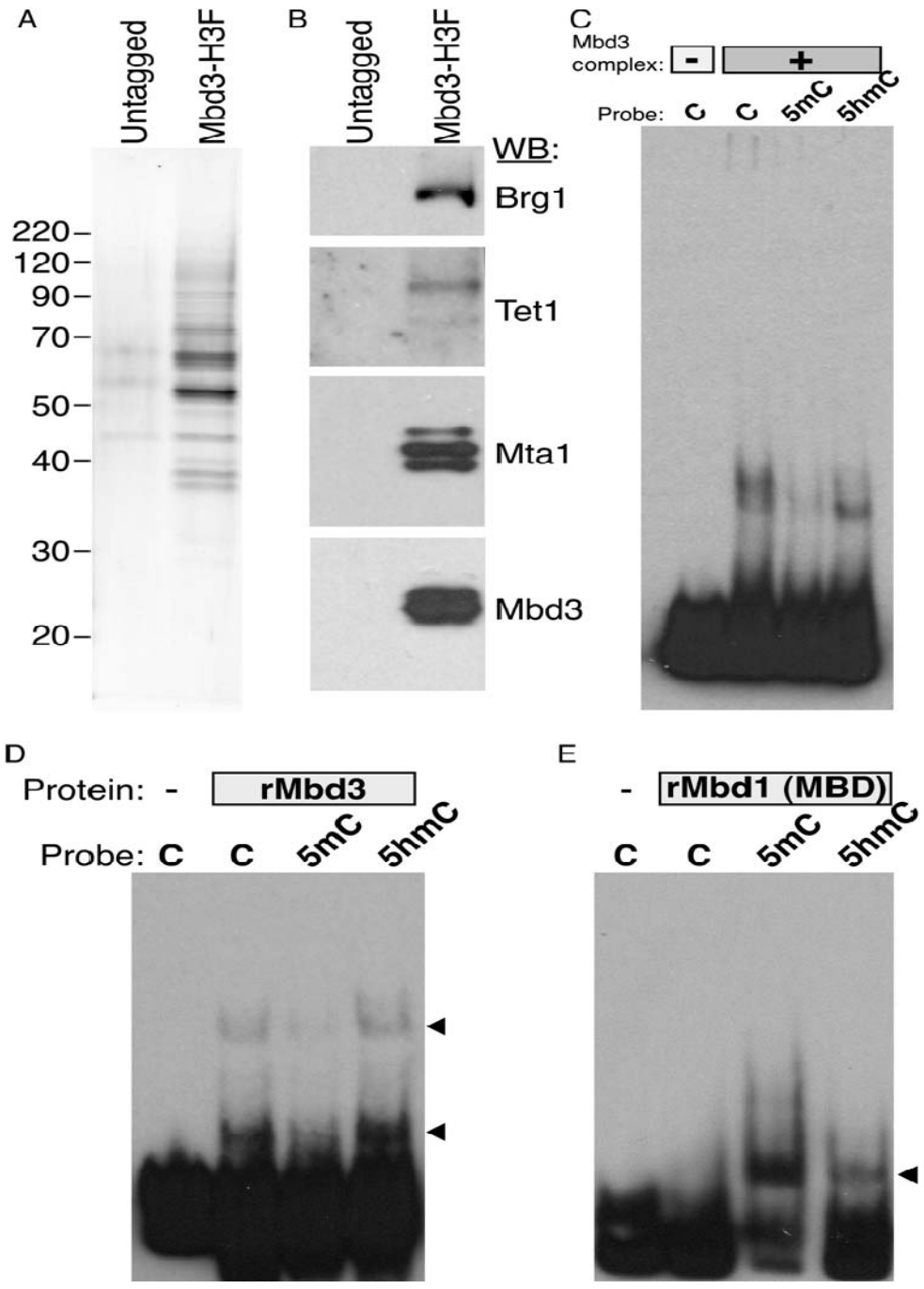

A
Untagged
Mbd3-H3F
220
120
90
70
50
40
30
20
B
Untagged
Mbd3-H3F
WB:
Brg1
Tet1
Mta1
Mbd3
C
Mbd3 complex: - +
Probe: C C 5mC 5hmC
D
Protein: - rMbd3
Probe: C C 5mC 5hmC
E
- rMbd1 (MBD)
C C 5mC 5hmC

Figure 2-6

Figure 2-6. Mbd3 directly binds to hydroxymethylated DNA *in vitro*.

(A) Silver stain showing tandem affinity purification from untagged or Mbd3-6His-3XFLAG ES cells.

(B) Western Blot of purifications described in (**A**) for indicated proteins. Mta1 is a component of the Mbd3/NURD complex. Preliminary mass spec results also identified most other known NURD subunits in this purification (not shown).

(C) EMSA assay using Mbd3/NURD complex and DNA probes containing unmodified cytosine (C), methylated cytosine (5mC) or hydroxymethylated cytosine (5hmC).

(D-E) EMSA assay using recombinant mouse Mbd3 (**D**) or recombinant Mbd1 methyl-binding domain (**E**), and various modified probes as in (**C**).

See also **Figure S2-6**.

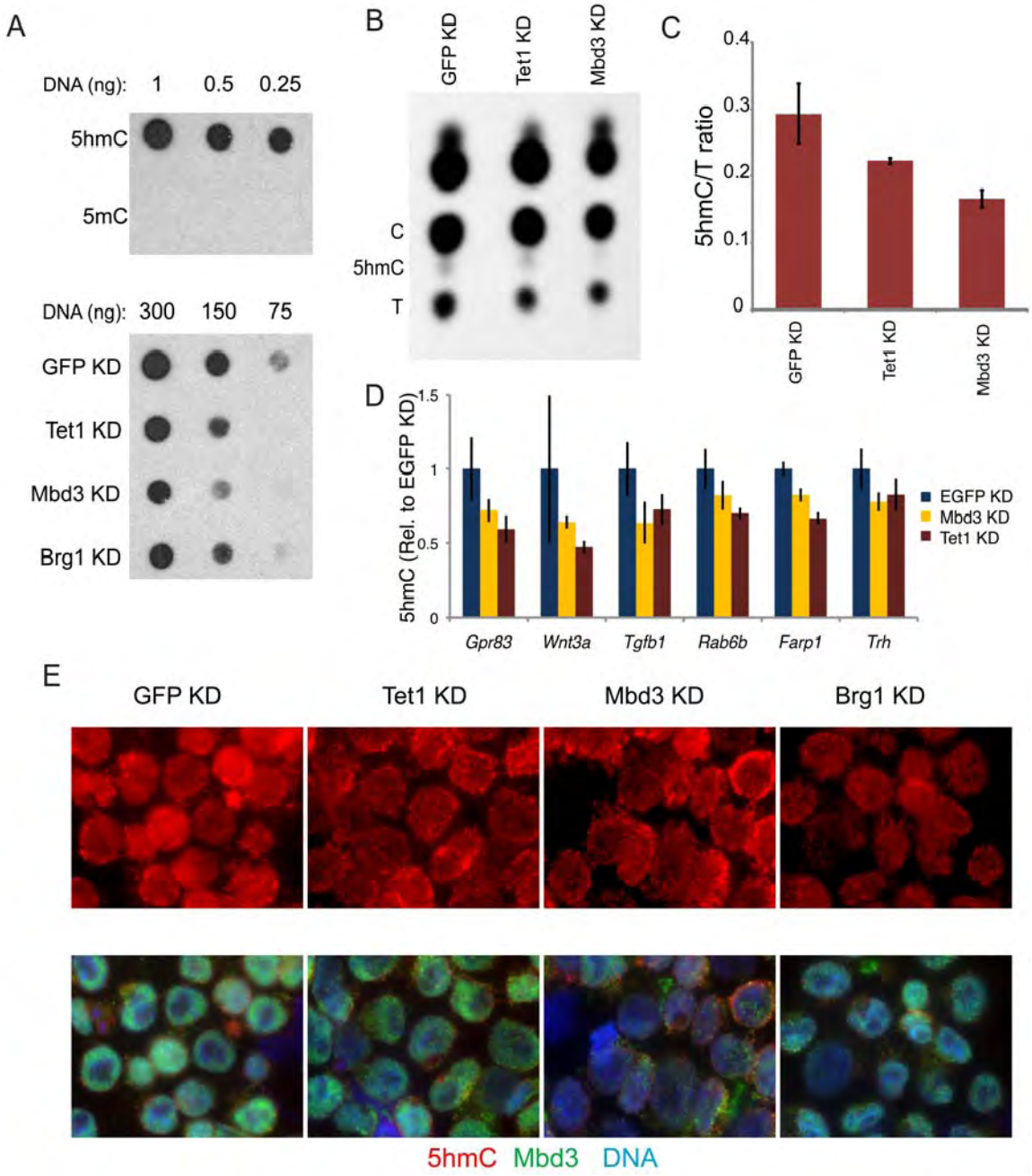

A
DNA (ng): 1 0.5 0.25
5hmC
5mC
DNA (ng): 300 150 75
GFP KD
Tet1 KD
Mbd3 KD
Brg1 KD
B
GFP KD
Tet1 KD
Mbd3 KD
C
5hmC
T
C
0.4
0.3
0.2
0.1
0
5hmC/T ratio
GFP KD
Tet1 KD
Mbd3 KD
D
5hmC (Rel. to EGFP KD)
1.5
1
0.5
0
Gpr83
Wnt3a
Tgfb1
Rab6b
Farp1
Trh
EGFP KD
Mbd3 KD
Tet1 KD
E
GFP KD
Tet1 KD
Mbd3 KD
Brg1 KD
5hmC Mbd3 DNA

Figure 2-7

Figure2-7. Mbd3 is required for global hydroxymethylation *in vivo*.

(A) Dot blots of 5hmC. Top panel shows positive (5hmC) and negative (5mC) controls for antibody specificity. Bottom panels show dilution series of genomic DNA isolated from the indicated knockdown ES cells. Mbd3 and Brg1 KDs have similar effects to Tet1KD on bulk 5hmC levels.

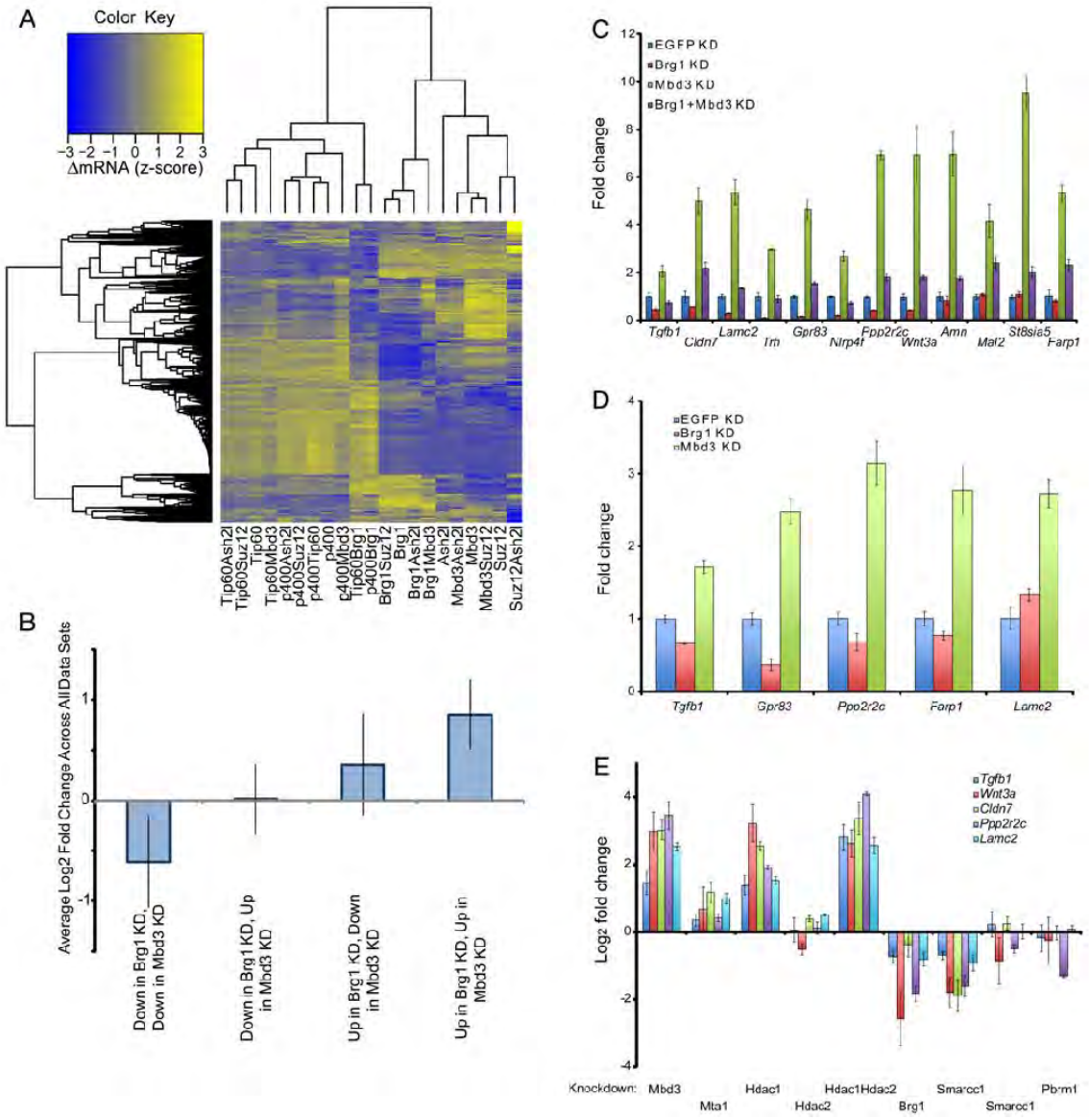
(B) Thin layer chromatography of radioactively end-labeled bases from MspI-digested genomic DNA (Ficz et al., 2011a) from the indicated knockdowns.

(C) Quantitation of 5hmC levels (normalized to levels of T) measured as in (**B**) for 5 independent replicate experiments. Columns show mean +/- sem.

(D) Tet1 and Mbd3 knockdowns have similar effects on hydroxymethylation of target gene promoters. Hydroxymethylated DNA was isolated by capture of biotin-glucosylated 5hmC-containing DNA fragments (Song et al., 2011), and fold enrichment over input was assessed by qPCR at the indicated loci and expressed as fold change relative to 5hmC levels in control (EGFP) KD cells. Data represent mean +/- sem.

(E) Immunofluorescence imaging of Mbd3 and 5hmC. Immunofluorescence images are pseudocolored blue (DAPI), green (Mbd3), and red (5hmC) for the indicated KDs – top panel shows 5hmC data only, bottom panel shows all 3 colors.

See also **Figure S2-7**.



Supplemental Figure 2-1

Supplemental Figure 2-1. Complete gene expression dataset.

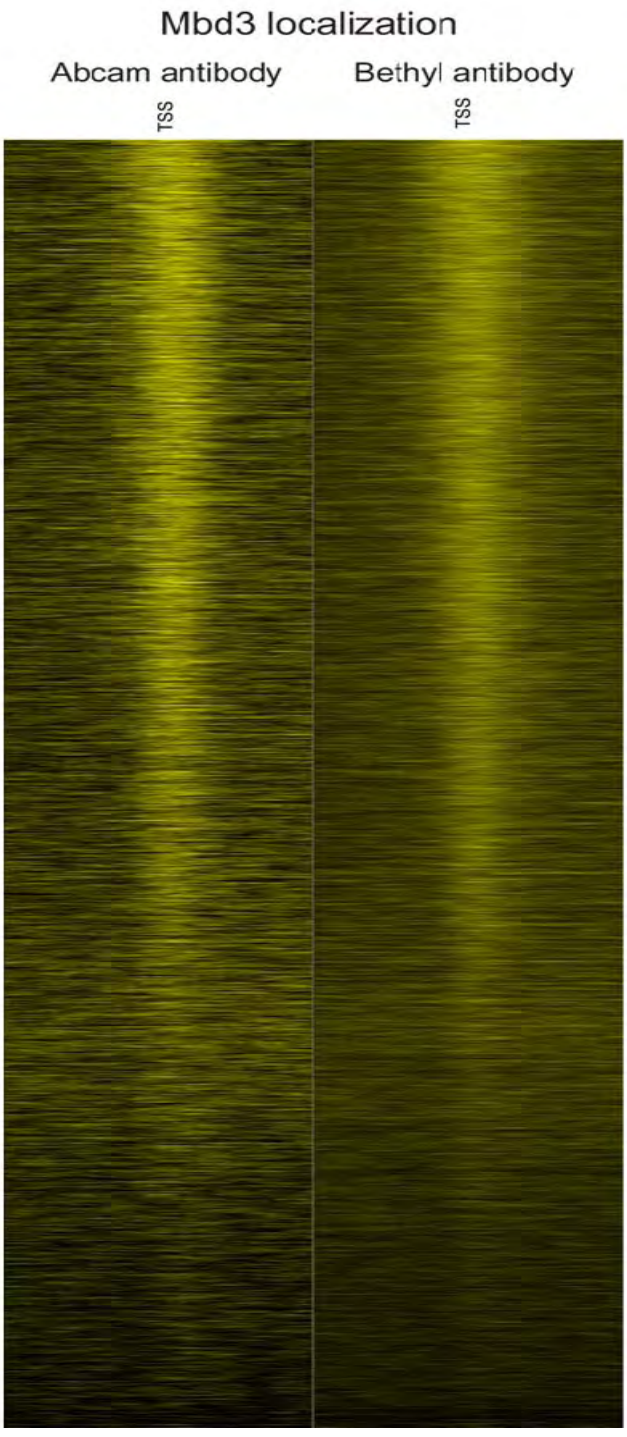
(A) Data from all 21 KD datasets (each representing the average of 2 replicates). A clustered heatmap of gene expression changes is shown. The most variant 300 genes that were significantly differentially expressed in each single KD were used for clustering.

(B) The Log2 (fold change) values of each gene significantly misregulated (adjusted p-value < 0.05) in both Mbd3 KD and Brg1 KD datasets were averaged across all 21 KD datasets (referred to below as “row averages”). Genes were separated into the indicated bins based on the direction of their expression changes in Brg1 KD and Mbd3 KD, and the mean and standard deviation of the row averages of the genes in each bin were calculated and plotted above. Notably, genes upregulated in both KDs were increased on average in all 21 KDs, and similar results hold for those downregulated in both KDs. Conversely, genes downregulated in Brg1 KD and upregulated in the Mbd3 KD show no general trend. P-values testing whether each bin is non-zero are indicated.

(C) Transcript levels of several target genes (downregulated in Brg1 KD and upregulated in Mbd3 KD) were assayed by RTqPCR in murine ES cells knocked down for the indicated regulators. Levels are shown relative to EGFP control KD. Levels are given as the average +/- sem of 3 technical replicates. Each experiment was performed at least two times with similar results.

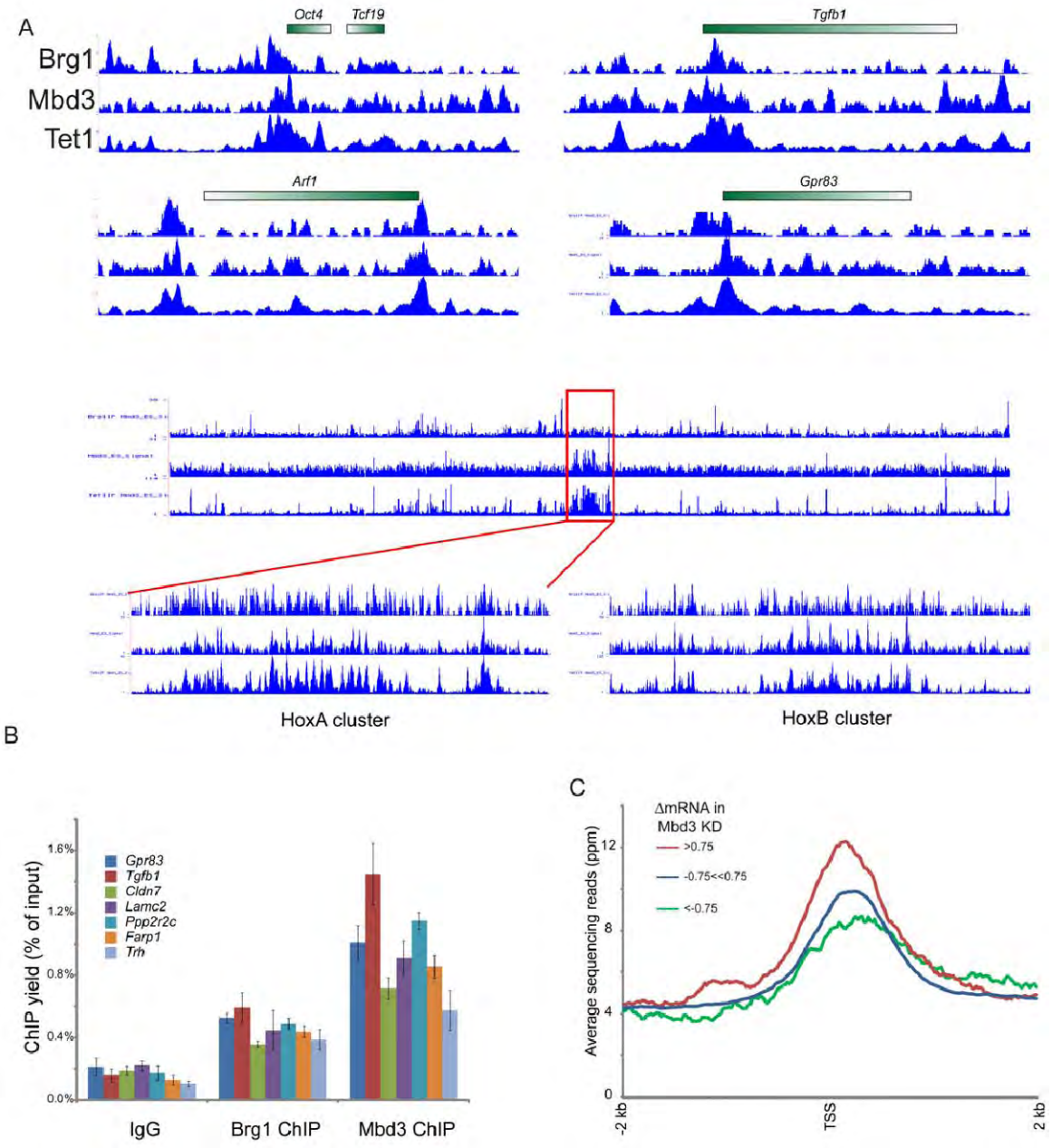
(D) As in (**C**), for Mbd3 and Brg1 knockdowns using a completely independent set of siRNAs. While KDs were slightly less efficient with this set of esiRNAs, the general effects of Brg1 and Mbd3 KD on target gene expression are maintained.

(E) As in (**C**), for knockdowns of additional NuRD (left) and BAF (right) subunits. Knockdown of NURD components Mta1, Hdac1, or double KD of Hdac1 and Hdac2 recapitulated Mbd3’s effects at five target genes. Similarly, knockdown of BAF subunit Smarcc1 recapitulated the effects of Brg1 knockdown on target genes, whereas knockdown of Polybromo, a subunit of the distinct PBAF complex, did not affect the genes assayed.



Supplemental Figure 2-2

Supplemental Figure 2-2. Mbd3 localization is robust to antibody source. TSS-aligned Mbd3 mapping datasets are shown for the main dataset in this manuscript (obtained using an antibody from Abcam), and for a second dataset (using an antibody from Bethyl Laboratories). Data are sorted by Tet1 binding as in **Figure 2-5**. Data from the Bethyl antibody had more sequencing reads but higher genome-wide background, and so almost all figures are derived from the Abcam dataset, with the exceptions of **Figures S2-3A** and **5B**, which use both datasets combined for the increased read number.



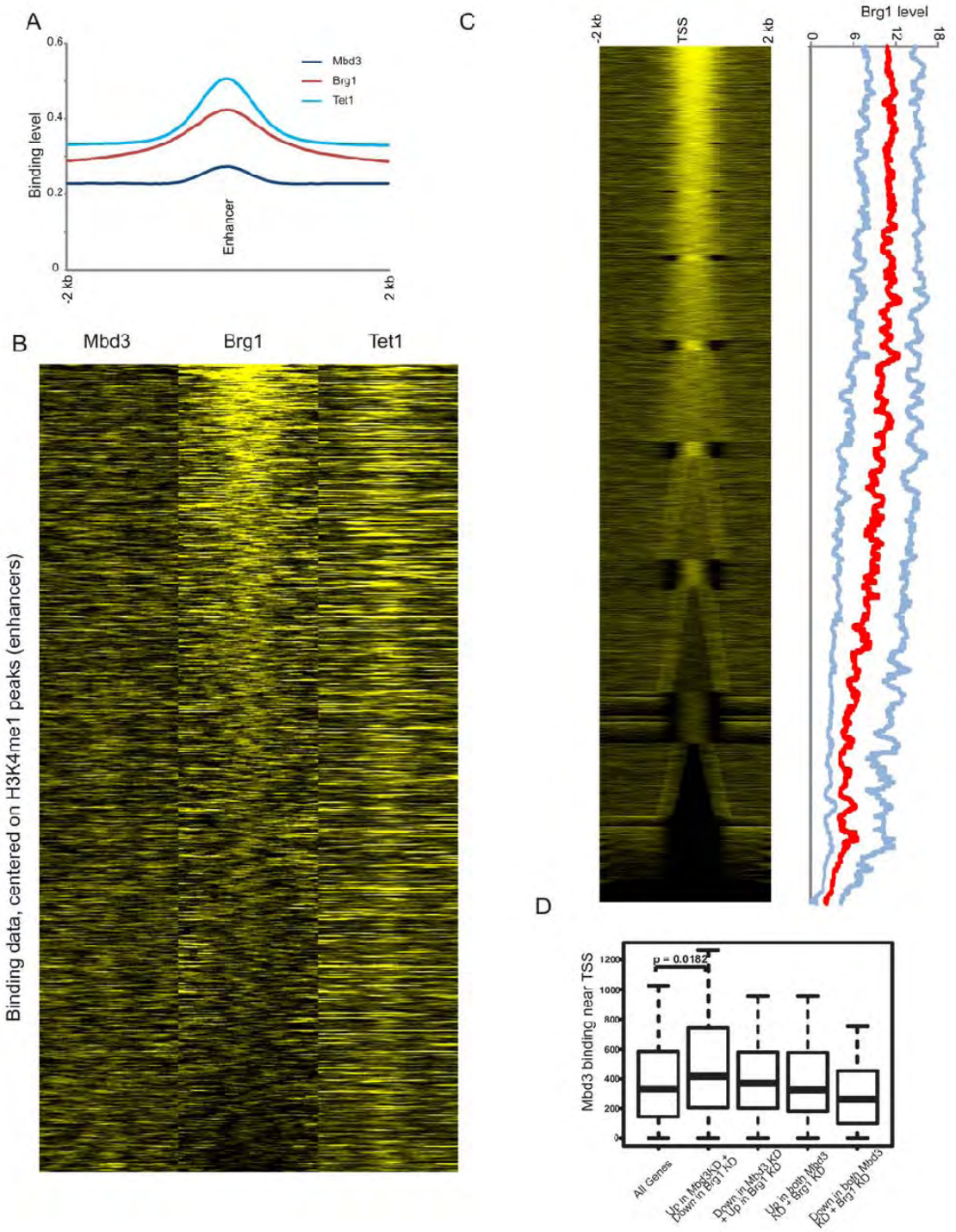
Supplemental Figure 2-3

Supplemental Figure 2-3. Genome-wide maps of Mbd3 binding in ES cells.

(A) Example genomic loci showing binding data for Mbd3 (this study), Brg1 (Ho et al., 2009a), and Tet1 (Wu et al., 2011c). The first four panels show regions surrounding individual genes of interest. The third panel down shows a long (chr6:50,300,000-54,010,000) region surrounding the HoxA gene cluster (boxed in red), which is then magnified as indicated. The HoxB cluster is shown to the right of the HoxA cluster, as indicated. Notably, Mbd3 is strongly associated with all 4 Hox loci (as shown here, and data not shown), with similar localization pattern to Tet1 (Wu et al., 2011c) and polycomb group proteins (Boyer et al., 2006).

(B) Sheared input samples and DNA immunoprecipitated with the indicated antibodies (or IgG from non-immunized rabbits) were subjected to quantitative PCR using promoter-proximal primers specific for the indicated genes. Shown is the enrichment, relative to input, obtained in each of the ChIP samples. Data shown are averages +/- sem of three technical replicates. Experiment was repeated twice with similar results.

(C) Mbd3-repressed genes are associated with Mbd3. Genes are grouped by the effects of Mbd3 KD as indicated, and averaged Mbd3 binding levels are plotted relative to gene's TSS. Among genes that could be mapped to TSSs with Mbd3 binding data, 250 genes were downregulated at this level in the Mbd3 KD (log2 change > 0.75), 17301 were unchanged, and 441 were upregulated.



Supplemental Figure 2-4

Supplemental Figure 2-4. Brg1 and Mbd3 bind to common targets near the TSS.

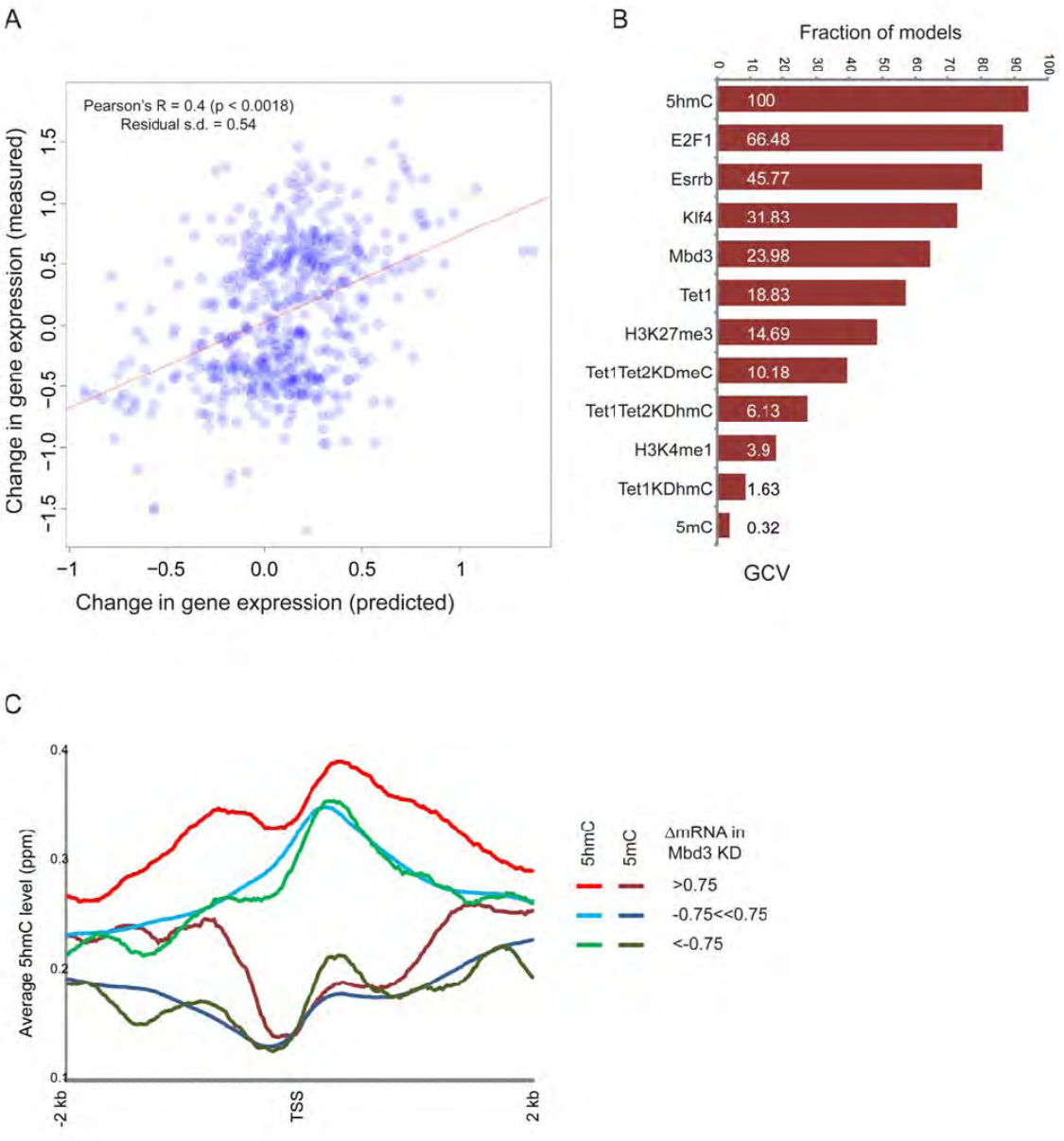
(A) Mbd3 is weakly associated with enhancers. Loci associated with the enhancer-specific histone mark H3K4me1 (Zentner et al., 2011) were used as anchors for mapping Mbd3, Brg1, and Tet1 mapping data. Averaged data for all enhancers are plotted.

(B) As in (**A**), but here a heatmap view shows all 75,772 putative enhancers, sorted by the extent of Brg1 binding to the enhancer. Brg1 and Tet1 both show strong localization to enhancers (and their localization is somewhat correlated), whereas Mbd3 levels are rarely above background at enhancers.

(C) Mbd3 and Brg1 colocalize at promoters *in vivo*. Mbd3 mapping data are sorted by the extent of Mbd3 binding to the 1 kb surrounding the TSS, and shown in the left heatmap. Brg1 binding to the 1 kb surrounding the TSS is shown on the right, with median binding shown in red and the bottom and top quartiles shown in grey/blue.

(D) Distribution of Mbd3 binding near promoters of genes regulated by both Mbd3 and Brg1. Shown are boxplots quantifying Mbd3 enrichment by ChIP-seq for genes misregulated in Mbd3 KD and Brg1 KD data sets. Genes were separated into the indicated categories based on the direction of the gene expression changes upon Mbd3 KD and Brg1 KD. Mbd3 enrichment was calculated by quantifying normalized read count within 1 kb in each direction from the transcription start site. Note that only one category, genes upregulated upon Mbd3 KD and downregulated upon Brg1 KD, was significantly enriched for Mbd3 binding compared to average.

Figure S5



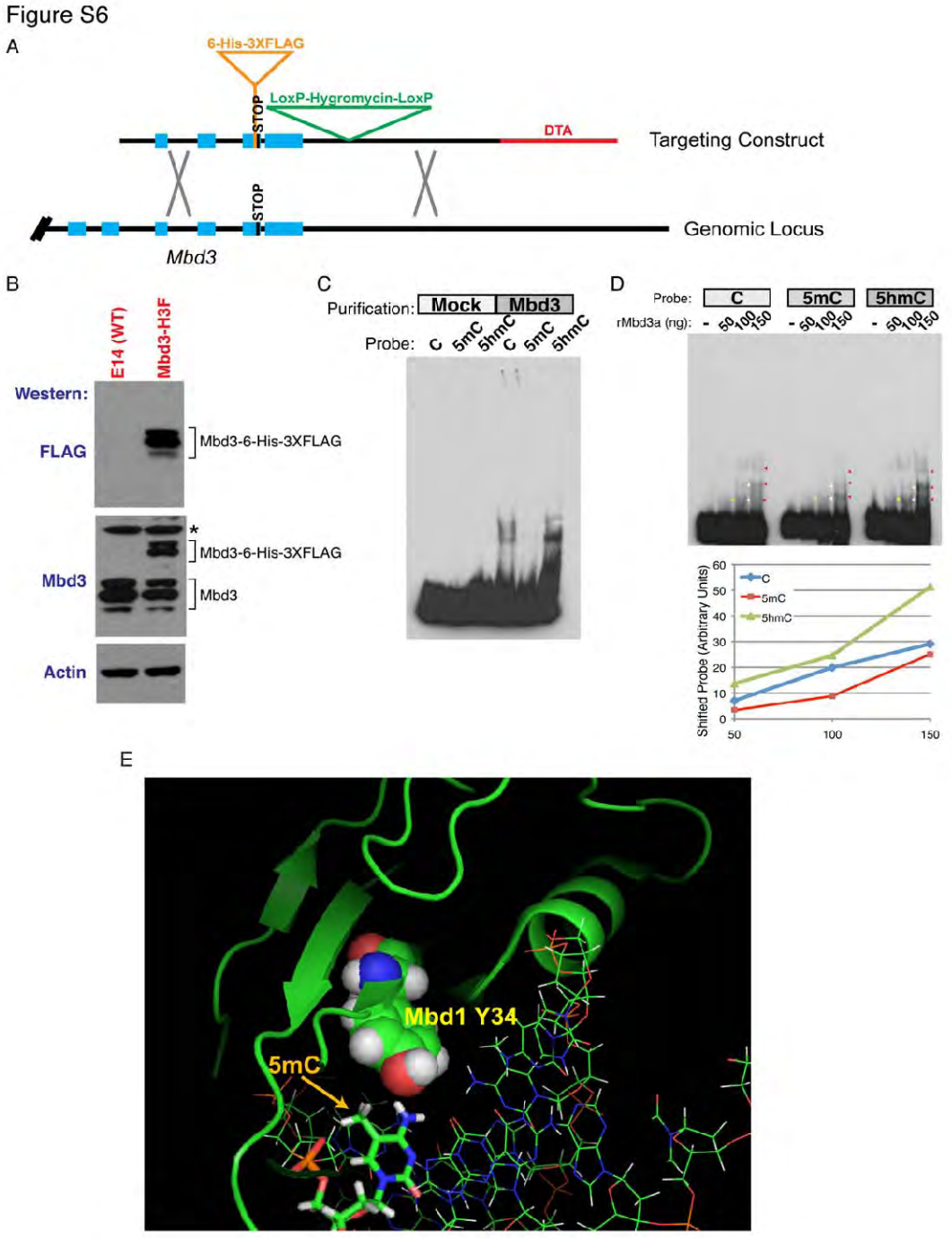
Supplemental Figure 2-5

Supplemental Figure 2-5. Multivariate prediction of Mbd3 function.

(A) Scatterplot showing good fit between a multivariate linear regression model predicting Mbd3 KD gene expression changes (x axis), and measured data (y axis). Line shows x=y. Here, model is used to predict genes exhibiting antagonistic regulation by Mbd3 and Brg1. Importantly, a highly similar model also significantly captures measured data for all 19,602 annotated genes (R=0.13; p-value < 2.2e-16), but because of the noise introduced by ~18,000 genes that are unaffected by Mbd3 KD (either unbound by Mbd3 or unaffected by Mbd3 binding due to redundant regulators) we focused on the model presented.

(B) Major contributors to the model. Top twelve localization datasets (features) contributing to the multivariate model (R=0.4 at 10 fold cross validation; p-value < 0.0018) are shown as a bar graph, with each bar showing the fraction of MARS models that pick the feature during feature selection, and the white number on the bar showing the total decrease in GCV (generalized cross validation) upon removing the feature from all models that include the feature.

(C) Mbd3-repressed genes are associated with high levels of hydroxymethylated cytosine, but low levels of methylated cytosine. 5hmC and 5mC mapping data (Ficz et al., 2011a) are aligned by TSS, and averaged for the indicated sets of genes as defined by their mRNA abundance changes in Mbd3 KD.



Supplemental Figure 2-6

Supplemental Figure 2-6. Mbd3/NURD binding to DNA is inhibited by cytosine methylation.

(A) Strategy for generation of epitope-tagged Mbd3 ES cells. Shown are the 3' end of the Mbd3 gene plus downstream sequence (below) and the targeting construct for introducing the C-terminal tandem 6His-3XFLAG tag (above). See Supplemental Experimental Procedures for details of targeting.

(B) Expression of Mbd3 isoforms in untagged cells (E14) or cells tagged at one of the two copies of Mbd3, as in (**A**). Note that the epitope tag induces a molecular weight shift in all three Mbd3 isoforms expressed in ES cells, but all three isoforms are expressed at their normal levels. Actin is shown as a loading control.

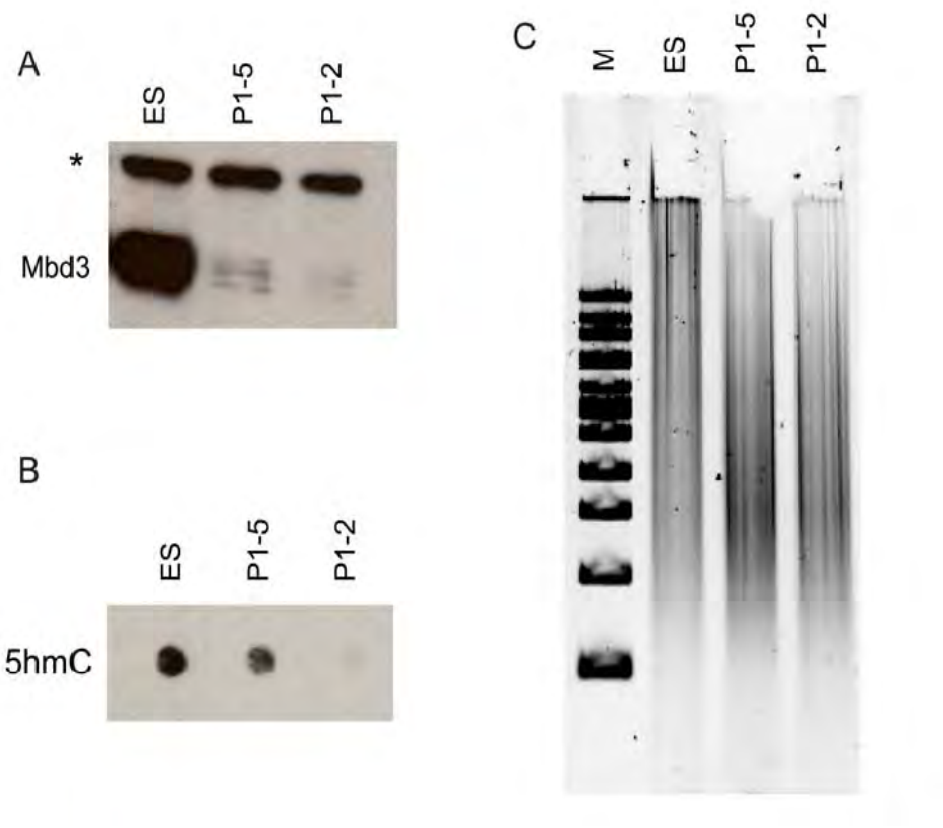
(C) Hydroxymethylated probe is not shifted by nonspecific contaminants. Extracts from untagged ES cells were subject to an identical purification scheme as those from tagged cells, and isolated material was subjected to EMSA assays. Shown is a side-by-side comparison of Mock purified and Mbd3 purified complexes bound to each of the probes from **Figure 2-6**.

(D) (Above) Titration of recombinant Mbd3 in gelshift experiments using unmodified, methylated, and hydroxymethylated probes. Arrows (compare like colored arrows for each probe) indicate multiple Mbd3-shifted bands, likely corresponding to binding of multiple Mbd3 proteins to the relatively long DNA probe used for the gelshift. **(Below)** Image J was used to quantify the total amount of shifted probe (plotted on y axis) at each titration for each different probe.

(E) Structural rationale for Mbd3 binding to 5hmC. Crystal structure (PDB accession 1IG4) of human Mbd1 complexed with methylcytosine-containing DNA (Ohki et al., 2001) is shown. Tyrosine 34 on Mbd1 is shown as spheres, while one of the two 5-methyl cytosines is shown as sticks, as indicated. Intriguingly, human/mouse Mbd3 differ from the Mbd1 methyl binding domain at very few residues, including a substitution of phenylalanine at position 34 in place of Mbd1's tyrosine. The loss of the hydroxyl group caused by this substitution may

be required to enable binding to DNA carrying an additional hydroxyl group on methylcytosine.

Figure S7



Supplemental Figure 2-7

Supplemental Figure 2-7. Mbd3 affects bulk 5hmC levels.

(A) Characterization of stable Mbd3 KD cells. ES cells stably expressing two different shRNAs directed against Mbd3 were generated as described in the Supplemental Experimental Procedures. Western blot shows Mbd3 protein levels in standard ES cells and in two different stable KDs. Asterisk shows background band on Western.

(B) Dotblot for 5hmC for the three cell lines from (**A**). 300 ng of DNA from each cell line are probed here. Note that the relatively high levels of 5hmC in line P1-5 likely results both from incomplete knockdown of Mbd3 (**A**), and from increased DNA loading (below).

(C) Loading control for (**B**). Identical amount of genomic DNA as loaded in (**B**) is shown here on an ethidium bromide-stained agarose gel.

CHAPTER 3

Histone turnover is conserved from yeast to mice and different histone variants have different turnover profiles in mouse embryonic stem cells

ABSTRACT

Nucleosome de-stabilization is important for all DNA templated processes. Nucleosomal positioning and stability particularly affect the transcriptional regulation of underlying DNA. Recent studies suggest that histone turnover not only plays a role in accessibility of the regulatory DNA elements but also provides dynamic alternating chromatin states by replacing old nucleosomes with newly synthesized variants. Here we investigate the histone turnover phenomenon in mouse embryonic stem cells and show that it is conserved from yeast to mammals. We show different histone variant dynamics at specific loci. Moreover, we generate animals expressing inducible constructs of different histone variants, which will enable us to extend histone turnover studies to whole animals and allow us to study histone dynamics at different developmental stages and in different tissues *in vivo*.

INTRODUCTION:

Eukaryotic genomes are organized into chromatin through packaging with highly conserved basic proteins called histone proteins. The histone octamer, consisting of 4 pairs of 4 different histone proteins – H2A, H2B, H3 and H4 – is wrapped by approximately 147 bp of DNA duplex (Luger et al., 1997b) and creates the fundamental unit of chromatin called the nucleosome. Packed nucleosome arrays can be folded into higher order chromatin structures in the nucleus in association with the linker histone H1 (Marino-Ramirez et al., 2005). Both the protein composition of the histone octamer (with canonical or variant

histones) and post translational modifications of these histones as well as nucleosome positioning dictate the overall organization of chromatin that influences all DNA templated processes including transcription, replication and DNA repair.

One of the most exciting aspects of chromatin is its potential to carry non DNA-based heritable information, known as epigenetic information (Henikoff et al., 2004; Rando and Ahmad, 2007). Evidence that chromatin carries epigenetic information comes from many fields, but most notably it comes from the observation that many epigenetic inheritance paradigms genetically require chromatin regulators. Cell state inheritance in metazoans requires Polycomb and Trithorax group chromatin regulators, whereas epigenetic inheritance in budding yeast requires the Sir complex, a chromatin regulating complex. Moreover, Polycomb has been shown to remain associated with replicating DNA in vitro, demonstrating that chromatin regulators can be inherited through replication in the absence of additional cellular machinery (Francis et al., 2009).

Mammalian development is characterized by a progressive loss of pluripotency, as stem and progenitor cells differentiate into specific cell types. Establishment and maintenance of cellular identity requires mechanisms that control specific sets of genes. These mechanisms utilize gene activation and/or silencing when and where they are needed (Orlando, 2003). These differential gene expression patterns are different in different types of cells with identical DNA sequence, underscoring the non-genetic mechanisms responsible for genomic plasticity (Theise and Wilmut, 2003). These considerations make

embryonic stem cells a key subject for epigenetic research, as these pluripotent cells can generate a wide range of distinct epigenetic states that can then self-propagate in the absence of inducing signals (Loebel et al., 2003; Nishikawa et al., 2007).

Over the past decade many studies have aimed to identify the factors that allow this plasticity. In order to answer this question many of those focused on ES cell specific transcriptional networks (Boyer et al., 2005; Wang et al., 2006). Although tissue specific transcription factors are very important in defining cellular identity, chromatin related factors also play important roles in maintaining and stabilizing it. While a defined set of transcription factors are sufficient to reprogram differentiated cells into embryonic stem cell state (Takahashi and Yamanaka, 2006; Wernig et al., 2007) or to another differentiated cell state (Zhou et al., 2008), reprogrammed cells nonetheless show preference to re-differentiate into their cell type of origin (Hu et al., 2010), and they may carry residual DNA methylation signatures of their originating cells (Kim et al., 2010b; Kim et al., 2011). Moreover, perturbation of chromatin regulators such as DOT1L, Suv39h1 (Onder et al., 2012), and Brg1 (Singhal et al., 2010) affects the efficiency and kinetics of cell reprogramming, indicating a role for chromatin factors in maintenance and disruption of epigenetic states. Not only are chromatin regulators involved in reprogramming of differentiated cells to the pluripotent state, but a multitude of chromatin factors are also involved in the process of differentiation. Genetic ablation of factors ranging from Polycomb to the TIP60 complex to components of the NURD complex and others have all been shown

to have defects in either the maintenance of the pluripotent state, or defects in differentiation (Boyer et al., 2006; Creyghton et al., 2008; Fazzio et al., 2008b; Kaji et al., 2006).

Thus, genetic dissection of the pluripotent state increasingly implicates chromatin regulators as major factors in controlling pluripotency, and this has motivated a large number of analyses of the chromatin state of ES cells. Relative to that of most differentiated cell types, embryonic stem cell chromatin possesses several unusual properties. Histone methylation patterns typical of both PcG and TrxG – repressive and active marks respectively – co-exist at promoters of developmentally important transcription factors (“master regulators”) in embryonic stem cells. Moreover, these so called ‘bivalent domains’ carrying both active and repressive histone modifications largely resolve upon differentiation (Bernstein et al., 2006). Thus, this ambivalent state of chromatin may have a role in ES cell plasticity by providing a flexible arrangement in the organization of the pluripotent genome. Another unusual characteristic of chromatin structure in embryonic stem cells has been revealed by FRAP (Fluorescence Recovery after Photobleaching) experiments. Although the inner core of nucleosomes, i.e the histone H3 and H4 tetramer, is very stable in differentiated cells (Kimura and Cook, 2001), FRAP studies showed that histones have a “hyper dynamic” nature in ES cell chromatin; they present high fluorescence recovery rate suggesting a high histone turnover (Meshorer et al., 2006). Compellingly, the fluorescence recovery rates substantially decrease upon differentiation to neural precursor

cells suggesting a role for histone dynamics in the maintenance of genomic plasticity.

Metazoan genomes encode multiple isoforms of the histone proteins, not only encoding the canonical histone proteins but also a set of replication independent variants (Jenuwein and Allis, 2001; Turner, 2000). Histone variants are localized to distinct chromatin domains, and they play roles in wide variety of processes (Talbert and Henikoff, 2010), suggesting that histone variants may also play important roles in other chromatin-based processes such as control of the pluripotent state in embryonic stem cells.

Two versions of canonical histone H3 are encoded in mammals, and are deposited during genomic replication – H3.1 and H3.2. In addition, mammals carry three replication-independent H3 variants. These include the centromere specific H3 CenpA, the sperm specific H3t (Witt et al., 1996), and H3.3, which is expressed throughout the cell cycle and maintained during differentiation (Ahmad and Henikoff, 2002; Wu and Gilbert, 1981). In humans, mice and flies H3.3 is expressed from 2 different genes, H3.3A and H3.3B. Both genes encode the same protein sequence, but have distinct promoter elements (Frank et al., 2003) and untranslated regions (Bramlage et al., 1997), suggesting differential transcriptional and post-transcriptional regulation. H3.3 protein was found to be associated with transcriptionally active genes and regulatory elements (Ahmad and Henikoff, 2002; Goldberg et al., 2010), as well as transcriptional elongation (Wirbelauer et al., 2005). In line with these findings, H3.3 was found to be enriched for active chromatin marks (Hake et al., 2006; McKittrick et al., 2004;

Sutcliffe et al., 2009). Its distinct distribution in the genome (Drane et al., 2010; Goldberg et al., 2010; Mito et al., 2005) and importance especially for certain cell types (Orsi et al., 2009; Sakai et al., 2009; Torres-Padilla et al., 2007) suggest that H3.3 is more than just a surrogate for H3.1 and it is functionally important for normal chromatin regulation.

While the microscopy studies described above for embryonic stem cells by necessity measure whole-genome histone dynamics, histone dynamics have been studied at nucleosomal resolution in other cell types and other organisms (Deal et al., 2010; Dion et al., 2007; Goldberg et al., 2010; Mito et al., 2005, 2007; Rufiange et al., 2007) . Both in yeast and fly, histone turnover is non-uniform across the genome, and rapid turnover coincides with specific DNA elements, such as chromatin boundaries or Polycomb/Trithorax response elements (Deal et al., 2010; Dion et al., 2007; Mito et al., 2007). Findings in those studies, notably rapid histone turnover at promoter regions of actively transcribed genes, suggest that histone turnover may have an important role in gene regulation, as higher histone turnover rates may provide higher probability of DNA element exposure. Also intriguing is a connection between histone replacement and epigenetic inheritance – epigenetically-regulated loci such as homeotic genes in flies or subtelomeric genes in yeast are associated with stably-bound nucleosomes, whereas the boundaries to epigenetic domains are associated with rapid turnover. Despite the proposed models, not much is known about the mechanism of action or the biological consequences of hyper dynamic chromatin states.

We therefore set out to measure histone dynamics in embryonic stem cells across the mouse genome. Previous studies in yeast utilized genetic “pulse chase” systems for induction of epitope-tagged histones to follow incorporation of newly-generated histones into chromatin over time course experiments. We chose to adapt this strategy to embryonic stem cells, and generated cell lines carrying inducible epitope-tagged histone variants in order to measure dynamic incorporation of H3.3 and of MacroH2A2 genome-wide in ES cells. Our results extend prior studies of H3.3 mapping in ES cells to a dynamic regime, and reveal completely novel facets of the biology of the H2A variant macroH2A. Finally, we have generated mice from the various ES cell lines, enabling future studies on tissue-specific histone dynamics in a multi-cellular organism.

RESULTS AND DISCUSSION

In order to assay histone variant dynamics genome-wide in embryonic stem cells and cell types derived from them, we generated ES cells based on the murine ES line KH2 (Beard et al., 2006) carrying a genomically-integrated cassette to enable tetracycline-inducible expression of HA-tagged histone variants (**Figure 3-1A**) confirmed by southern blots (**Figure 3-1B**). Cell lines were generated for several different histone variants, including canonical H2A, macroH2A2 and H3.3 as well as stable cell line for inducible knockdown of macroH2A1.2. But here we detail results from two – MacroH2A2 (hereafter called Macro), and H3.3. **Figure 3-2A** shows Western blots for HA-H3.3 and HA-Macro expression over a time course of doxycycline induction. In both cell lines, epitope-tagged histone expression was detectable by 2-3 hours, and increased

over ~12 hours. Even after 12 hours of induction, ectopically expressed tagged histone levels were orders of magnitude lower than the endogenous proteins levels; barely detectable at 12hrs with respect to endogenous H3 (**Figure 3-2B red arrow**). However, this did not perturb the key biology of ES cells, as even after 72 hours of overexpression, ES colonies did not differentiate as assessed by cell morphology, continued to express pluripotency markers such as Oct4, Sox2 and Nanog (**Figure 1C**), and stained positively for alkaline phosphatase (**Figure 3-1D**) (See also **Supp Figure3-1** for colony staining for Oct4).

Genome-wide dynamics of H3.3 replacement in mES cells

Because metazoans encode distinct histone H3 variants for replication-coupled deposition (H3.1) and replication-independent replacement (H3.3) (Ahmad and Henikoff, 2002), we sought to determine whether the pulse-chase approach that is necessary in budding yeast (Dion et al., 2007; Rufiange et al., 2007) provided any additional insights into histone dynamics. In other words, if steady-state mapping of H3.3 localization is sufficient to identify sites of histone replacement genome-wide, what additional information is gained by explicitly including a pulse chase component to such mapping experiments?

We therefore focus first on the dynamics of H3.3 replacement, the well-studied H3 variant involved in replication-independent nucleosome replacement. ES cells carrying tetracycline-inducible HA-H3.3 were treated with doxycycline for 0, 3, or 6 hours, and HA-H3.3-containing chromatin was mapped genome-wide by chromatin immunoprecipitation followed by Illumina deep sequencing (ChIP-Seq). Sequencing reads were mapped back to the genome. Importantly, HA

mapping at t=0 (no doxycycline) did not show enrichment over specific loci but rather genome-wide nonspecific background, demonstrating the specificity of the anti-HA antibody.

Because H3.3 replacement is strongly associated with the 5' ends of genes (Mito et al., 2005, 2007), we aligned all annotated genes by their transcription start sites (TSSs), and averaged all mapped reads at each position relative to the TSS (**Figure 3-3A**). Consistent with studies in flies and mouse embryonic stem cells (Goldberg et al., 2010; Mito et al., 2005, 2007), we find that H3.3 is localized to two peaks surrounding the TSS. Also consistent with published reports, we found that H3.3 levels correlate with transcription rate of the associated gene (**Figure 3-3B**). Our results with H3.3 localization at the 6 hour time point are therefore broadly consistent with known aspects of H3.3 localization.

Importantly, we also observe additional details of H3.3 localization not observed in steady-state mapping studies. Specifically, at the 6 hour time point, as in Goldberg et al, H3.3 peaks surrounding the TSS of highly-transcribed genes are of similar heights, suggesting that nucleosome turnover upstream and downstream of the TSS occurs at similar rates. However, at earlier time points the upstream peak is lower. We interpret these results as demonstrating that H3.3 replacement is more rapid downstream of the TSS than upstream, a result that is consistent with the more rapid downstream nucleosome replacement measured by a metabolic labeling method (Deal et al., 2010) and consistent with the greater number of short transcripts generated downstream of promoters

relative to upstream (Core and Lis, 2008). Moreover, H3.3 dynamics observed at polycomb bound regulatory elements in flies are also present in mouse embryonic stem cell genome at regions occupied by polycomb proteins Rnf2 and Suz12 (**Figure 3-3C**).

These results imply that under steady-state mapping conditions (e.g. Goldberg et al), and after extended induction in a pulse chase system, nucleosomes exhibiting moderate to high turnover rates become saturated with H3.3. Thus, even for the canonical turnover variant, we find that kinetic studies provide additional information on histone dynamics.

Specific histone chaperones are responsible for the differential localization of H3 variants. The canonical H3s, H3.1 and H3.2, associate with CAF-1 that travels behind the replication fork with the proliferating cell nuclear antigen (PCNA) and assembles into chromatin in a replication coupled manner (Shibahara and Stillman, 1999). While in yeast the only H3 variant associates both with Caf1 and HirA, in vertebrate HIRA binds only to replication independent H3 variant H3.3 (Tagami et al., 2004). The amino acid substitutions in the histone fold domain in H3.3 allow this selective association with replication independent chromatin assembly factors (Ahmad and Henikoff, 2002). Recent studies showed that HIRA is not the only chaperone for H3.3 assembly. In fact, various chaperones are responsible for the differential targeting of H3.3 throughout mammalian chromatin. While HIRA is responsible for H3.3 deposition at promoter regions and transcribed regions of active genes; Daxx-Atrx complex is

required for H3.3 localization at pericentric heterochromatin and telomeres (Drane et al., 2010; Lewis et al., 2010; Wong et al., 2010).

Additional comparison of our data set with static mapping in the presence and absence of different H3.3 chaperones suggested that most of the dynamics are result of neither HIRA nor Daxx/Atrx function (**Figure 3-3D**). Moreover, enrichment of high turnover rates at active promoters and enhancers indicate the existence of a separate H3.3 chaperone responsible for the H3.3 dynamics at regulatory regions. Interestingly, DEK is another novel H3 chaperone which has been shown to have H3.3 specificity in vitro and is potentially responsible for H3.3 deposition at regulatory elements (Sawatsubashi et al., 2010);

Genome-wide dynamics of MacroH2A replacement

We next extended our studies to a histone variant with unknown dynamic properties, macroH2A. MacroH2A (also known as H2AFY) is unique among histone variants in that it has a large, distinct C-terminal domain, which includes a small stretch of basic amino acids and a putative leucine zipper motif. While the N-terminal domain of the protein shares 65% amino acid sequence similarity with the canonical H2A, the C-terminal domain (also known as the macro domain) is almost twice as big as the canonical H2A and has no relation to other histones (Chadwick and Willard, 2001a). The macroH2A genes are located on different chromosomes: macroH2A1 on 13 and macroH2A2 (also known as H2AFY2) on chromosome 10. The two genes have nevertheless the same architectural genomic organization with conserved exon-intron boundaries except for the splicing variant site in macroH2A1. MacroH2A1 encodes two splicing variants

macroH2A1.1 or macroH2A1.2; both are conserved in mammals. MacroH2A2 is 80% identical to macroH2A1.2 and it is 98% conserved between mice and human (Chadwick and Willard, 2001a; Costanzi and Pehrson, 2001; Zhang et al., 2005). Although they are preferentially expressed at different levels in different cell types in a tissue, two proteins seem to share overall similar genomic localization; for instance, they both localize to the inactivated X chromosome (Xi). The enrichment of macroH2A on the Xi chromosome of female mammalian cells was interpreted as evidence for its role in silencing (Costanzi and Pehrson, 1998), even though the association of macroH2a with the Xi appears to occur after the random inactivation of the second X chromosome (Chadwick and Willard, 2001a). Moreover, macroH2a localization to Xi depends on the non coding RNA XIST but is not required for the inactive state. When XIST is conditionally deleted, gene silencing is maintained despite the loss of macroH2a on Xi (Csankovszki et al., 1999). It is evident that macroH2a function extends beyond its role in MCB formation in Xi since it localizes to other chromatin domains and autosomal chromosomes. In addition, some of the organisms that do not undergo X inactivation still express this H2A variant (Abbott et al., 2005).

Since MacroH2A has not been extensively studied and has not been characterized dynamically, we therefore also examined MacroH2A dynamics using our pulse-chase system. We mapped MacroH2A genome-wide at 3, 6, and 12 hours after induction of the HA-tagged variant by ChIP-Seq as for H3.3. Reads were mapped back to the genome and genes were aligned by TSS as above.

Because MacroH2A localization has not been characterized in embryonic stem cells, we chose unbiased K means clustering for our initial analysis of MacroH2A localization patterns. We found K=4 yielded the most informative distinctions between clusters (**Figure 3-4A**). One cluster was associated with background levels throughout the time course and will not be discussed further. Two clusters were comprised of genes exhibiting increasing levels of MacroH2A throughout the time course (see also Figure 3B), with MacroH2A being found in domains several kb long either upstream (Cluster 2) or downstream (Cluster 3) of the TSS. Finally, we found a surprising cluster of genes with very high levels of MacroH2A surrounding TSSs early in the time course (**Figure 3-4A**). The tagged MacroH2A appears to become gradually depleted at later time points which we believe due to the normalization method used for deep sequencing data sets (See below).

These results are consistent with two populations of MacroH2A-containing chromatin that can be distinguished by their dynamic behavior. We infer that the TSS-proximal MacroH2A that is enriched at early time points before being lost represents a rapidly-exchanging population of MacroH2A that is present at low steady-state occupancy, while the larger blocks that are gradually gained correspond to higher-occupancy MacroH2A domains undergoing either slow turnover or replication-coupled assembly. Because deep sequencing datasets are normalized to account for sequencing depth, enrichment of the rapidly-exchanging population is high at early time points before assembly of HA-MacroH2A into the slower subpopulations, whereas at later time points

normalization relative to the widespread MacroH2A in cold domains results in a diminishing TSS-proximal peak.

MacroH2A2 generally localizes TSS-proximally in ES cells and MEFs

Our results showing rapid TSS-proximal turnover of macroH2A were quite surprising, given published results showing that MacroH2A1 localizes in broader domains covering silent genes. This raised the concern that we might artifactually be driving inappropriate MacroH2A into promoters as a result of overexpression of this variant. We therefore carried out genome-wide localization of endogenous MacroH2A.2, in wild type cells not bearing the tagged variant, using ChIP-Seq with an antibody against the MacroH2A2 protein. As shown in **Figure 3-4**, MacroH2A was strongly localized to a peak surrounding the TSS in embryonic stem cells and in embryonic fibroblasts. Thus, our findings of TSS-proximal macroH2A turnover do not reflect artifactual incorporation during overexpression.

We next sought to compare Macro localization to turnover. To our surprise, only a fraction of promoters with steady-state MacroH2A peaks were subject to rapid MacroH2A replacement, with roughly half of the TSS-proximal peaks showing no incorporation of HA-MacroH2A even after 12 hours of induction (**Figure 3-5A**). We sought characteristics of the genes with promoter peaks of MacroH2A with and without high rates of turnover, finding that those promoters exhibiting rapid MacroH2A turnover were generally expressed at much higher levels than those genes with MacroH2A peaks that were not dynamic (**Figure 3-5B**). Taken together, our data show that MacroH2A.2 is widely localized to a large number of promoters in both pluripotent cells and in

differentiated cells, but that it is dynamic at only a subset of its binding sites. We are currently repeating these experiments for publication, and we are investigating the functional consequences of the difference in MacroH2A.2 dynamics by assaying gene expression defects in knockout ES cells lacking MacroH2A.2.

Generation of turnover mice

To date, the majority of studies on histone dynamics have been carried out in cell culture systems. However, it will be of great interest in the future to understand tissue-specific differences in chromatin dynamics, both under control conditions and in response to environmental perturbations. Because embryonic stem cells are the cell type used for generating knockout and transgenic mice, we decided to generate mice carrying the tetracycline-inducible epitope-tagged histone variants.

HA-H3.3 ES cells were injected into Black6 X DBA F2 generation blastocysts, and embryos were implanted into pseudopregnant females. We obtained multiple chimeric animals with high levels of contribution from the ES strain background, and found after mating that several of these animals had germline contribution from the tet-inducible ES cells (**Figure 3-6A**).

As a proof of principle to survey tissue-dependent expression of the epitope-tagged H3.3, transgenic tet-HA-H3.3 mice were fed doxycycline (in drinking water) for several days. **Figure 3-6B** shows Western blots against HA for mice with or without doxycycline feeding. HA-H3.3 induction is observed in nuclear protein extracts from liver samples of animals that are fed with

doxycycline for 4 or 8 days but not in un-induced animals. These animals will therefore provide a unique and exciting resource for characterization of histone dynamics in different tissues from an animal.

METHODS

Targeting & Stable Cell Line Derivation

The unique EcoRI restriction site of the pBS31 vector used in order to integrate the Kozak sequence followed by HA tag and BspH1 sequences. cDNA of various histone variants ((H2a-MMM1013-98478233; H2Az-MMM1013-9498090; MacroH2a2- MMM1013-9201250; H3.3-MMM1013-98478016, H1o-MMM1013-65296, Open Biosystems & human H3.1-Kind gift of Eric Campeau)) was cloned into pBS31 vector using the unique BspH1 restriction site. This allowed the ectopic expression of histone variants with HA tag. Construct integrity was confirmed by western blot showing the inducible expression of HA-tagged histone variants. pBS31 expression/targeting vector contains a PGK promoter followed by an ATG start codon, an Frt site, and a splice acceptor-double poly A cassette. Tetracycline response element is at the upstream of CMV minimal promoter. The minimal promoter is followed by an SV40 poly-A signal at the downstream of the unique EcoRI cloning site. When pBS31 vector is electroporated into KH2 embryonic stem cells along with a Flpe recombinase expressing vector pCAGGS, it is integrated into the Collagen1a1 locus at the Frt site. In the mean time, Frt-flanked PGK-neomycin resistance cassette is targeted to the upstream of the promoterless, ATG-less hygromycin resistance cassette

located at downstream of the Frt site in the Col1A locus. KH2 embryonic stem cells also harbor the reverse trans-activator of tetracycline response element (M2-rtTA) at the ROSA26 locus and express it under the control of the ROSA26 promoter (Beard et al., 2006) . Together, the components of this system enable tetracycline induction of the epitope-tagged histone variant of choice in embryonic stem cells from a genomically-integrated construct. Hygromycin selection was started a day after the electroporation of HA-tagged variant constructs. Twenty days of selection of hygromycin resistance yielded numerous colonies, which were verified for proper flip-in to the Coll1a1 locus by southern blotting.

Generation of Tet-inducible MacroH2A1.2 KD Mice:

Potent shRNAs were designed to knockdown macroH2A1.2 using RNAxs web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAxs). Five different shRNAs, designed by the web server, were embedded in microRNA structure to allow Pol II derived expression of shRNAs as described by Zhou H. et al. (Zhou et al., 2005). Designed structure was flanked by EcoRI restriction sites and the resulting DNA sequence was synthesized by Epoch Biolabs. shRNAmir sequence was PCR amplified and cloned into the pBS31 targeting vector via the unique EcoRI site. Inducible for the macroH2a1.2 knockdown, stable embryonic stem cells were generated by successful genomic integration, which was confirmed by southern blot. Knockdown efficiency of different clones was tested by KD induction followed by both RT-qPCR and western blot (See

Supplementary Figure 3-1B). These embryonic stem cells were injected into Black6 X DBA F2 generation blastocysts to obtain inducible macroH2A KD mice.

Southern Blot and Confirmation of Induction:

Genomic DNA was isolated from selected colonies and SpeI-digested. Digestion of Col1a1PL plasmid with PstI/XbaI yielded the 850bp band which was gel purified and used as a 3' internal probe for the southern blotting. The wild type allele yielded a 6.2 kb band whereas the successfully flipped-in HA tagged histone variant inducible allele had a 4.1 kb band and a 6.7 kb band for the Frt-containing knock-in allele. Activation of the TetOn HA-tagged histone expression was carried out by addition of 2-5ug/mL doxycycline hyclate (Sigma D9891) to the culture media. Cells were collected at different induction time points and induction of HA tagged histone variants in ES cells observed via western blot.

Tris-Tricine PAGE and Western Blot

16% TrisTricine poly acrylamide gel was prepared with 30% (w/v) Acrylamide/Bis (4mL) and 3.3mL Tris.Cl-SDS solution (3M Tris base and 3% SDS pH 8.45), 1.3gr glycerol, fresh 10% (w/v) ammonium per sulfate (APS) and TEMED. Samples were resuspended in 2X Tricine sample buffer (100mM Tris.Cl, 200mm DTT, 0.02% Commassie Blue G250 and 0.1% SDS, pH 6.8) and run with anode (40mM Tris.Cl pH 8.9) and cathode (100mM TrisCl, 100mM Tricine, 0.1% SDS) buffers. PVDF membrane transfer was conducted in Tris-Glycine transfer buffer. The membrane was blocked with 10% milk and incubated with antibody (Beta Actin (ab8224), HA (ab9110), Pan H3 (ab1791)).

Immunofluorescence Staining of KH2 cells

Cells were washed with PBS and fixed with 4% paraformaldehyde in 200 ul PBS for 10 minutes at room temperature. After washed for twice, 5 minutes with PBS, they were permeabilized and blocked with 5% FBS and 0.1% Triton X-100 in PBS at room temperature for 15 minutes. Cells were then incubated with primary antibody diluted 1:100 with 1% FBS and 0.1% Triton X-100 in PBS for 30 minutes at room temperature. Next they were washed 3 times with PBS for 5 minutes and incubated with secondary antibody diluted 1:200 with 1% FBS and 0.1% Triton X-100 in PBS for 30 minutes at room temperature. During the first PBS wash DAPI was added and incubated for 5 minutes. Cells then washed 3 more times for 5 minutes with PBS (Starting from the secondary antibody addition all steps were performed in dark).

Cytospin & Immunofluorescence Staining of KH2 cells

ES cells were trypsinized and pre-plated in gelatin coated dishes and incubated at CO2 incubator for an hour in order to eliminate the feeder MEFs. Cells were washed with PBS and counted. $5X10^5$ to $1X10^6$ cells were resuspended in 200ul 4% formaldehyde in PBS, fixed for 30 minutes at room temperature. Fixed cells were then washed twice with PBS for 5 minutes at room temperature. After attachment of the cells on glass slides via cyto-spin, they were permeabilized and blocked with 5 %FBS and 0.1% TritonX-100 in PBS for 15 minutes at room temperature. Exact same protocol described above was applied to the slides.

Alkaline Phosphatase Staining

Formaldehyde-fixed ES colonies were incubated for 15 minutes at 37 °C with alkaline phosphatase (SK5100, Vector labs) which was diluted in 100mM Tris-HCl pH 8.5, pluripotent ES colonies stained red.

MEF Isolation from MacroH2A-HA Chimeras

Mouse embryos were isolated at embryonic day E12.5 followed by the removal of internal organs. Carcasses were minced in the presence of 0.25% Trypsin-EDTA and incubated at 37ºC for 20 minutes. MEF medium was added and cell suspension was titrated followed by plating cells onto 15 cm culture dishes, one to two dishes per embryo. Cells were cultured at 37°C, 3% CO_2, and 5% O_2 for 48 hours. The cells were then trypsinized and frozen.

Geneotyping for Inducible Mice

Rosa26 locus was amplified with cycling parameters of 95^{o} C for 45" (1 cycle); 95 ^{o} C for 45", 55 ^{o} C for 45", 72 ^{o} C for 1' (35cycles); 72 ^{o} C for 5' (1 cycle); 4 ^{o} C) with primer set of three (1: 5'AAAGTCGCTCTGAGTTGTTAT; 2: 5'GCGAAGAGTTTGTCCTCAACC; 3: 5'GGAGCGGGAGAAATGGATATG) for discrimination of wild type versus M2 reverse trans-activator (M2rtTA) integration. Wild type allele product size is 500 base pairs (bp), M2rtTA allele product size is 300bp. Col1A locus was amplified with touchdown PCR protocol with cyclin parameters of 95°C for 1' (1 cycle); 94°C for 30", 70°C for 45" (2cycles); 94°C for 30"; 68°C for 45" (5cycles); 94°C for 20", 66°C for 1' (29cycles); 4°C, with a

primer set of 3 (col/frt-B:CCCTCCATGTGTGACCAAGG; col/frtA1:GCACAGCATTGCGGACATGC; col/frtC1:GCAGAAGCGCGGCCGTCTGG) for the discrimination of wild type versus frt integration of tagged variants. While wild type allele product size is 331 bp, flip-in allele product size is 551 bp.

Time Course & ChIP-Seq

80% Confluent ES cells were pre-plated on gelatin and incubated for 45 min for the removal of the feeder MEFs. Then they were split in three plates each of which was induced in different time points by the addition of final 2ug/mL doxycycline hyclate. At the end of the doxycycline incubation, the cells were crosslinked with formaldehyde (Sigma) to a final concentration of 1% and incubated at room temperature for 10 min. Crosslinking was quenched with 125mM Glycine.

Cells were washed twice with ice cold PBS containing PMSF then collected in PBS and pelleted at 1000g for 5 min, at 4^{o}C. Cell pellets were either flash frozen in liquid nitrogen and stored in -80^{o}C, or continued with sonication. Pellets were resuspended in 270 µl SDS-Lysis Buffer (1% SDS, 10mM EDTA and 50mM Tris-Cl, pH 8.1.) including protease inhibitor complex (Sigma) and PMSF (Sigma). Samples were sonicated in Bioruptor (UCD-200) at high settings for 3 times 15 min- 30sec on/30sec off cycle- followed by 14000 rpm spin at 4^{o}C for 1 hour. Supernatants were transferred to a new tube and pellets were resuspended in 150 µl of lysis buffer and processed in Bioruptor at previous

settings for two more 15 min cycles. Supernatants were collected after 13000 rpm spin at 4^{o}C for 10 min and combined with supernatants from the previous centrifugation. Chromatin amount was quantified with Abs 260 reading. 50 µg magnetic proteinA beads (NEB) were washed twice with 600 µl 0.1M sodium phosphate (NaK) buffer (pH 8.0) and antibody-coupled in 200 µl NaK buffer for 30min at room temperature. Antibody-coupled beads were blocked with 5mg/ml BSA-PBS solution for 1 hour at 4^{o}C. 70 µg of chromatin for each immunoprecipitaton was diluted to 1:10 in IP-blocking buffer (5mg/ml BSA and 5µg/ml yeast tRNA in 0.01% SDS, 1.1% Triton X-100, 1.2mM EDTA, 16.7mM Tris-Cl (pH 8.1), 167mM NaCl) combined with antibody coupled magnetic beads and incubated at 4^{o}C overnight (up to 16 hours). Magnetic beads were washed once with RIPA buffer (0.1%SDS, 10mM Tris (pH 7.6), 1mM EDTA, 0.1% Na-Deoxycholate, 1% Triton X-100); 3 times with NaCl buffer (0.5M NaCl in RIPA); 3 times with LiCl buffer (0.3 M LiCl, 0.5% NP40, 0.5% Na-Deoxycholate in RIPA); twice with 0.2% TritonX-100 TE buffer and once with TE buffer (pH 8.0).

Washed beads were resuspended in 300 µl TE and incubated at 65 ^{o}C overnight with final concentration of 3% SDS, 1mg/ml proteinaseK (Sigma) and 0.5mg/ml glycogen (Ambion). Eluted material was transferred to a new tube and the beads were resuspended in 150 µl 0.5M NaCl-TE solution and incubated at 65^{o}C for an hour. Eluted materials were combined and PCI (Phenol-Chloroform-Isoamylalcohol) extracted using phase-lock tubes (Eppendorf). Ethanol precipitated ChIP DNA was treated with RNAse (Qiagen) for 2 hours and with

CIP for an hour at 37^{o}C (calf alkaline phosphatase NEB; in 1x NEB Buffer3, 0.25U/µl CIP, reaction clean up with Qiagen MiniElute spin columns)

Deep Sequencing Library Construction.

ChIP material was then gel-purified (100 to 800bp in size) from a 2% TAE agarose gel using the MiniElute columns QiaQuick (Qiagen). Gel-purified DNA fragments were blunt-ended and phosphorylated with the EPICENTRE End-it-Repair kit (1X buffer, 0.25mM dNTPs,1mM ATP, 1ul/50ul reaction of enzyme mix) for 1 hour at room temperature and cleaned up with Qiagen MiniElute spin columns. Adenosine nucleotide overhangs were added using EPICENTRE exo-Klenow for 45min at room temperature (with 0.2mM dATP). Illumina genome sequencing adaptors were then ligated using the EPICENTRE Fast-Link ligation kit: 11.5 µl A-tailed DNA eluted from a MinElute column was mixed with 1.5µl 10X ligation buffer, 0.75µl 10mMATP, 0.5 µl Illumina DNA adaptors and 1 µl ligase. The reaction was incubated for 1 hour at room temperature and subsequently supplemented with 7.5 µl water, 1µl 10X buffer, 0.5µl 10mM ATP and 1µl ligase, and incubated overnight at 16^{o}C.The ligation reaction was cleaned up with MinElute columns (with an additional wash step to eliminate all the excess adaptors) and the adaptor-ligated fragments were amplified by PCR as follows: 0.75 µl of each Illumina genomic DNA sequencing primers, 6µl 10xPfx buffer 1.8µl 10mM dNTPs, 1.2µl 50mM MgSO4 and 1µl Pfx DNA polymerase (Invitrogen) were added to 30µl DNA template in a 100ul reaction. The cycling parameters were: 1. 94^{o}C for 2’; 2. 94^{o}C for 15’’; 3. 65^{o}C for 1’; 4. 68^{o}C for 30’’; 5.

repeat from step 2 16 times; 6. 68^{o}C 5'. The PCR product (250 to 450bp in size) was gel purified from a 2% TAE agarose gel using the QiaQuick columns (Qiagen). Gel-purified fragments were finally precipitated with 300 mM sodium acetate and absolute ethanol. Pellets were then resuspended (25nM final concentration) in TE buffer and sent for SOLEXA sequencing at the UMass Worcester core deep sequencing facility.

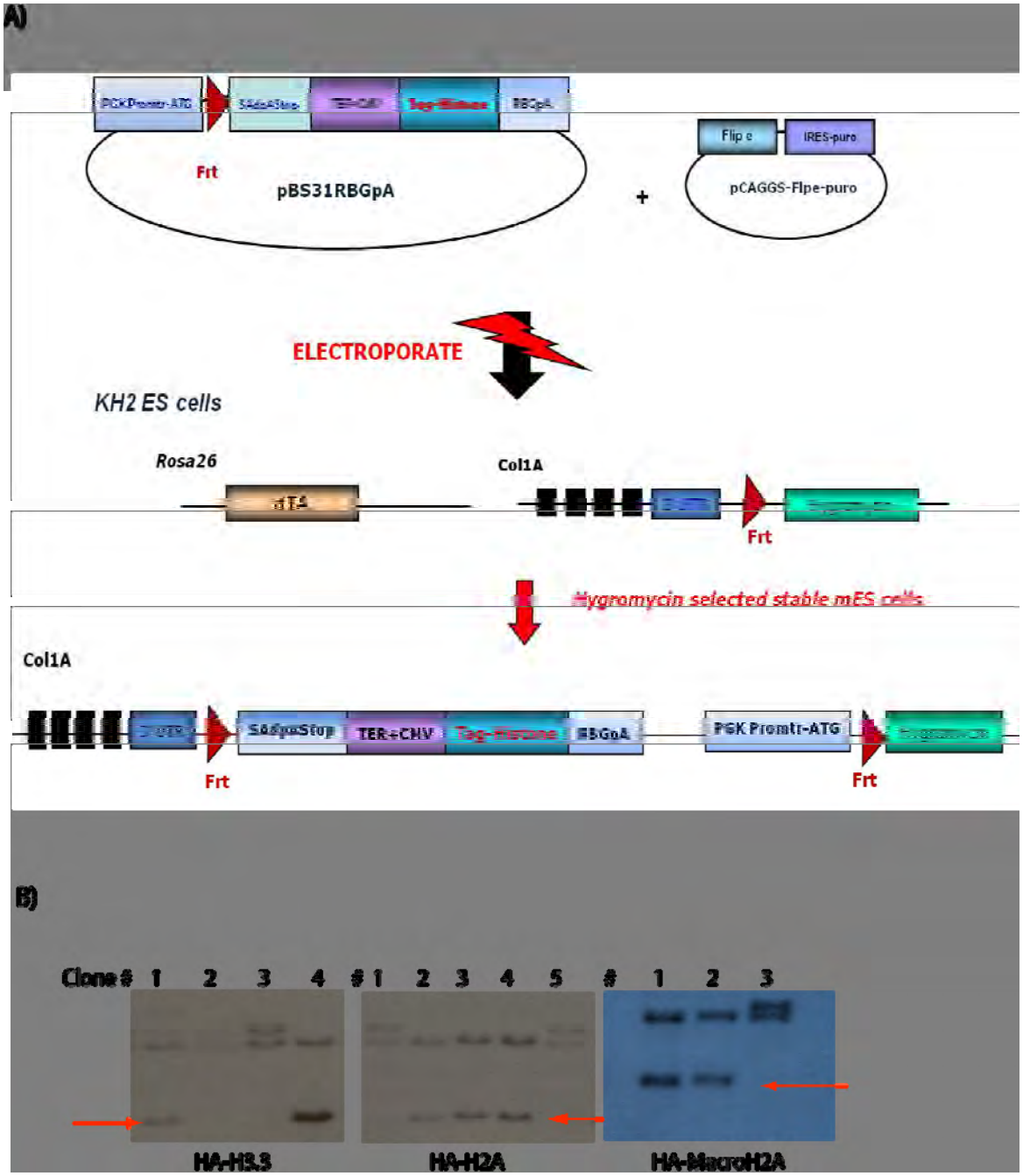

A)
Frt
pBS31RBGpA
+
Flip e
IRES-puro
pCAGGS-Flpe-puro
ELECTROPORATE
KH2 ES cells
Rosa26
Col1A
Frt
Hygromycin selected stable mES cells
Col1A
SA&polStop
TER+CHV
Tag-Histone
RBGpA
PGK Promtr-ATG
Frt
Frt
B)
Clone # 1 2 3 4 # 1 2 3 4 5 # 1 2 3
HA-H3.3
HA-H2A
HA-MacroH2A

Figure 3-1

Figure 3-1. Targeting strategy and conformation of genomic integration

A) The unique EcoRI restriction site downstream of tetracycline response element and CMV minimal promoter, used to integrate HA tagged histone variants into the pBS31 expression/targeting vector. HA-tagged variant integration site is followed by an SV40 poly A signal. The Frt site upstream of tetracycline response element allows the targeting of the construct downstream of Col1A locus in the presence of Flpe recombinase. In the mean time the PGK promoter followed by an ATG start codon gets integrated upstream of promoterless, ATG-less hygromycin resistance gene through the same Frt site. Genomic integration of the targeting construct is enabled by electroporation of KH2 embryonic stem cells with pBS31 vector along with a Flpe recombinase expressing vector pCAGGS. Correct integration of the construct renders KH2 cell resistant to hygromycin and allows for selection of stable cell lines. Expression of reverse trans-activator of tetracycline response element (M2-rtTA) under control of the ROSA26 promoter required for doxycycline responsiveness.

B) Southern blot performed on genomic DNA isolated from hygromycin resistant colonies. Correct integration of flipped in allele yield in two distinct bands, 6.7kb and 4.1kb (pointed out with red arrow).

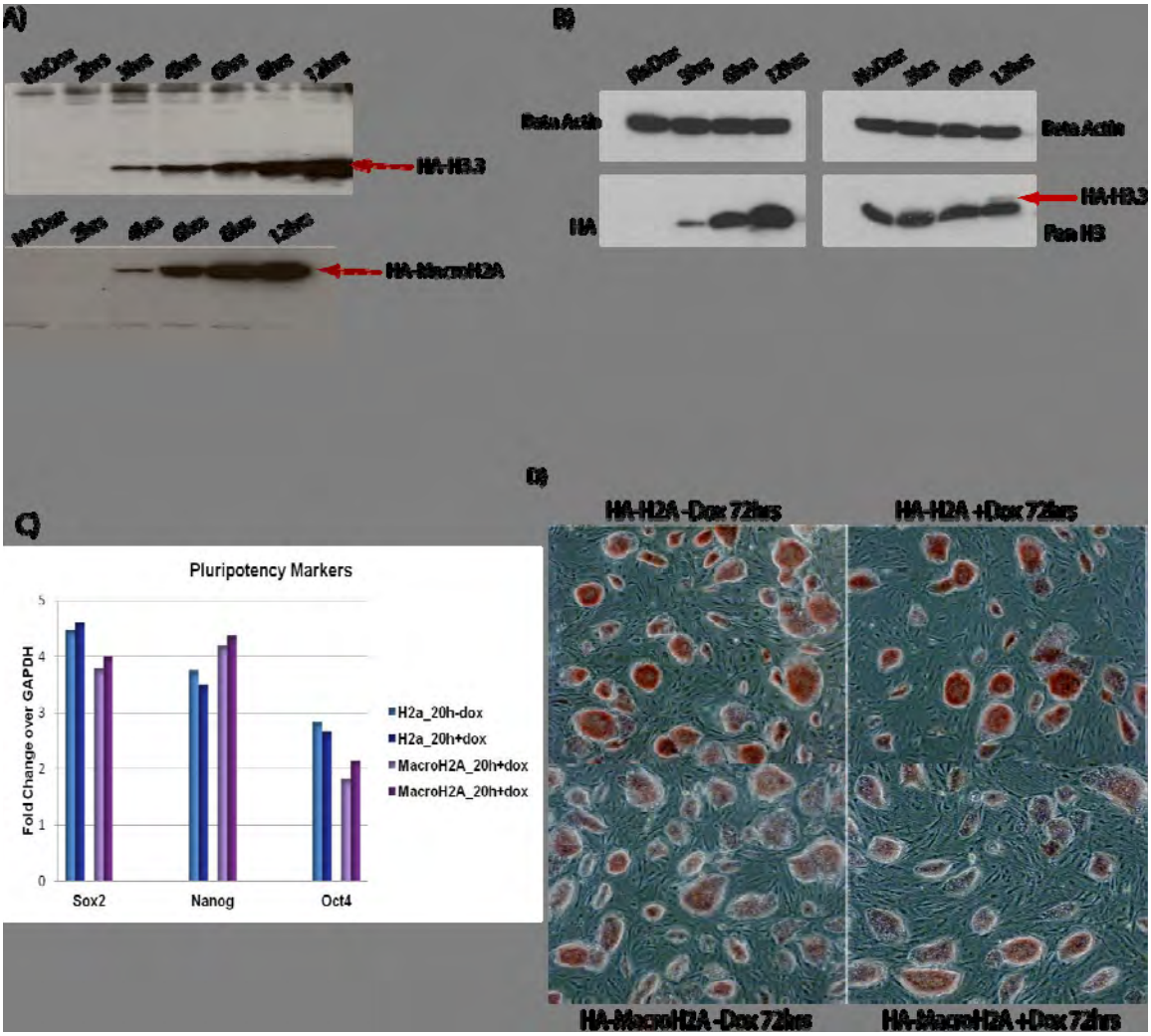

A)
HA-H3.3
HA-MacroH2A
B)
Beta Actin
Beta Actin
HA
HA-H3.3
Pan H3
C)
Pluripotency Markers
Fold Change over GAPDH
5
4
3
2
1
0
Sox2
Nanog
Oct4
H2a_20h-dox
H2a_20h+dox
MacroH2A_20h+dox
MacroH2A_20h+dox
D)
HA-H2A -Dox 72hrs
HA-H2A +Dox 72hrs
HA-MacroH2A -Dox 72hrs
HA-MacroH2A +Dox 72hrs

Figure 3-2

Figure 3-2. Ectopic expression of tagged histone variants does not affect pluripotency.

A) Time course induction of genomically integrated tagged histone variants. Upper blot shows HA tagged H3.3 and lower blot shows HA tagged MacroH2A expression blotted with HA antibody. No Dox and 2 hours of induction do not show HA tagged expression and only detectable after 3 hours of induction.

B) 16% Tris-Tricine gel was used to separate HA tagged H3.3 (red arrow) from endogenous H3 proteins and blotted with pan H3 antibody (blot at the far right bottom); showing that ectopically expressed tagged H3.3 level is far less than endogenous proteins.

C) RT-qPCR performed on total RNA extracted from cells induced or un induced (20 hrs) for expression of tagged histone variants and we did not observe change in the levels of core pluripotency factors.

D) ES colonies are stained for alkaline phosphatase and didn’t show difference in alkaline phosphatase levels after 72hrs induction compared to un-induced counterparts.

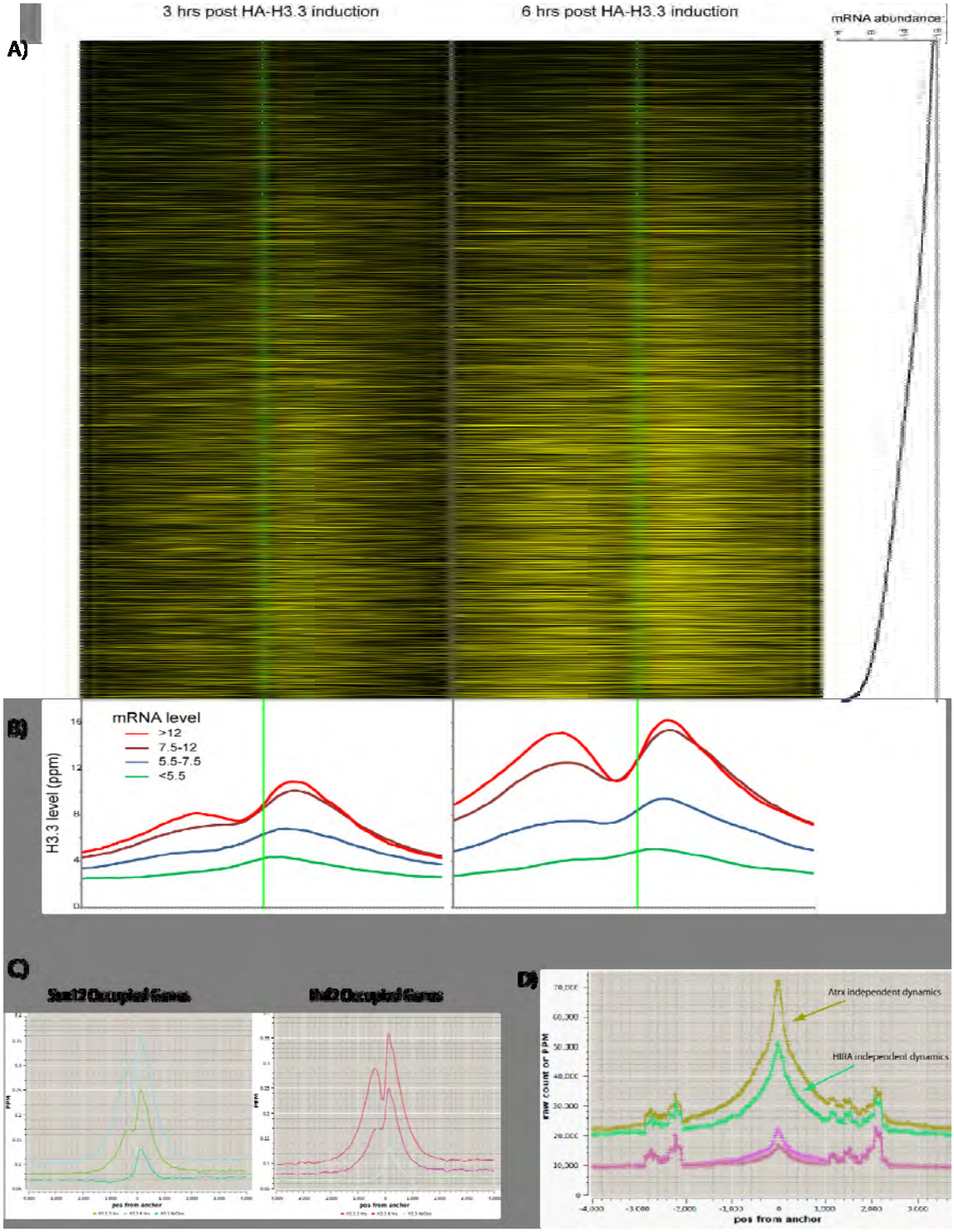

A)
3 hrs post HA-H3.3 induction
6 hrs post HA-H3.3 induction
mRNA abundance
B)
mRNA level
>12
7.5-12
5.5-7.5
<5.5
H3.3 level (ppm)
C)
D)
Atrx independent dynamics
HIRA independent dynamics
raw count or PPM
pos from anchor

Figure 3-3

Figure 3-3. H3.3 turnover is similar to H3.3 turnover observed in flies.

A) HA tagged H3.3 cells are induced for several hours and chromatin immune precipitated with HA antibody and deep sequenced. Heat map on the right shows HA tagged H3.3 enrichment at 6 hours post induction while the one on the left shows enrichment in 3 hours post induction. Genes are ordered by their expression levels and aligned by the transcription start sites (TSS is signified by the green line).

B) H3.3 enrichment is most prominent in highly expressed genes (red line) at downstream of TSS after 3hrs of induction (Graph on the left). After 6 hours of induction (graph on the right), nucleosomes both upstream and downstream of TSS are saturated with H3.3 in highly and moderately (red and maroon lines respectively) expressed genes.

C) H3.3 turnover also correlates with polycomb protein bound genes exemplified by Suz12 and Rnf2 occupied genes.

D) H3.3 dynamics at 6 hours post induction in H3.3 occupied loci of Atrx null cells signifies Atrx independent dynamics (fair green line). H3.3 dynamics at 6 hours of induction in H3.3 occupied loci of HIRA null cells signifies HIRA independent dynamics (lower bright green line).

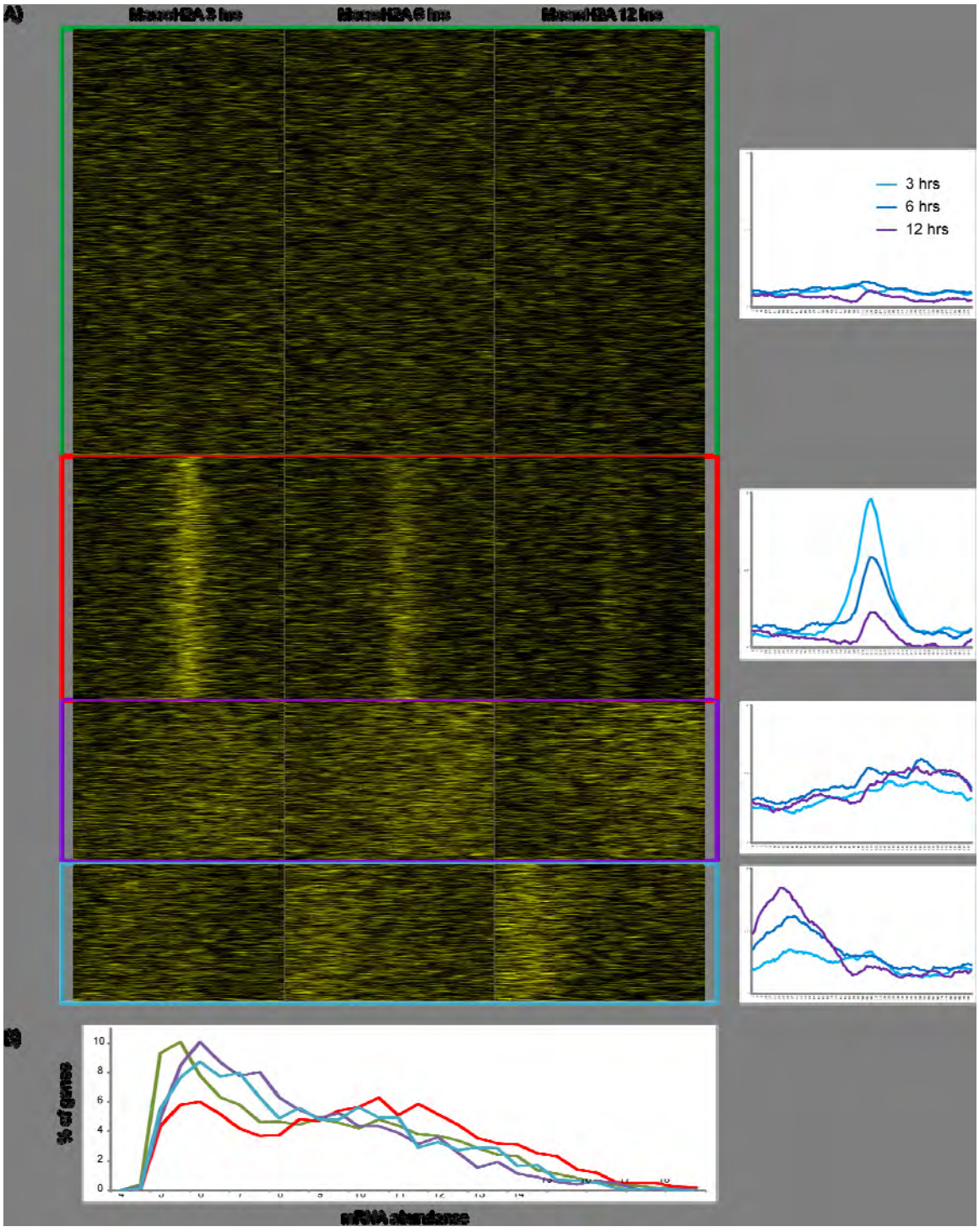

A)
MacroH2A 3 hrs
MacroH2A 6 hrs
MacroH2A 12 hrs
3 hrs
6 hrs
12 hrs
B)
% of genes
mRNA abundance

Figure 3-4

Figure 3-4. MacroH2A shows different turnover patterns and dynamics at differentially expressed genes.

A) Unbiased K means clustering of 3, 6 and 12hr induced tagged MacroH2A ChIP-Seq experiments that are aligned by TSSs yielded in 4 different clusters with distinct turnover rates and localization with respect to transcription start sites. First cluster (in green frame) shows genes that show no enrichment of tagged variant incorporation even in later time points, suggesting no or very low MacroH2A turnover. Cluster2 (in red frame) shows very rapid histone turnover evident by the high enrichment at early time point (at 3hr) and the appearance of gradual reduction in MacroH2A enrichment is due to normalization artifact because of this high turnover rate at early time points stays steady while slower turnover regions gets enriched MacroH2A in later time points. Clusters 3 and 4 (purple and blue frames respectively) show slow turnover rate, with gradual enrichment of MacroH2A at later time points (6hr and 12hrs) at downstream and upstream of TSS respectively. Meta gene analysis (at the right hand side of the heat map) summarizes turnover rates of each cluster at different time points with respect to TSS.

B) Expression levels (x-axis) plotted against the percentage of genes (y-axis) in each cluster (Cluster1, 2, 3 and 4; as green, red, purple and blue respectively). High turnover rate, observed in cluster 2 corresponds highly expressed genes' transcription start site.

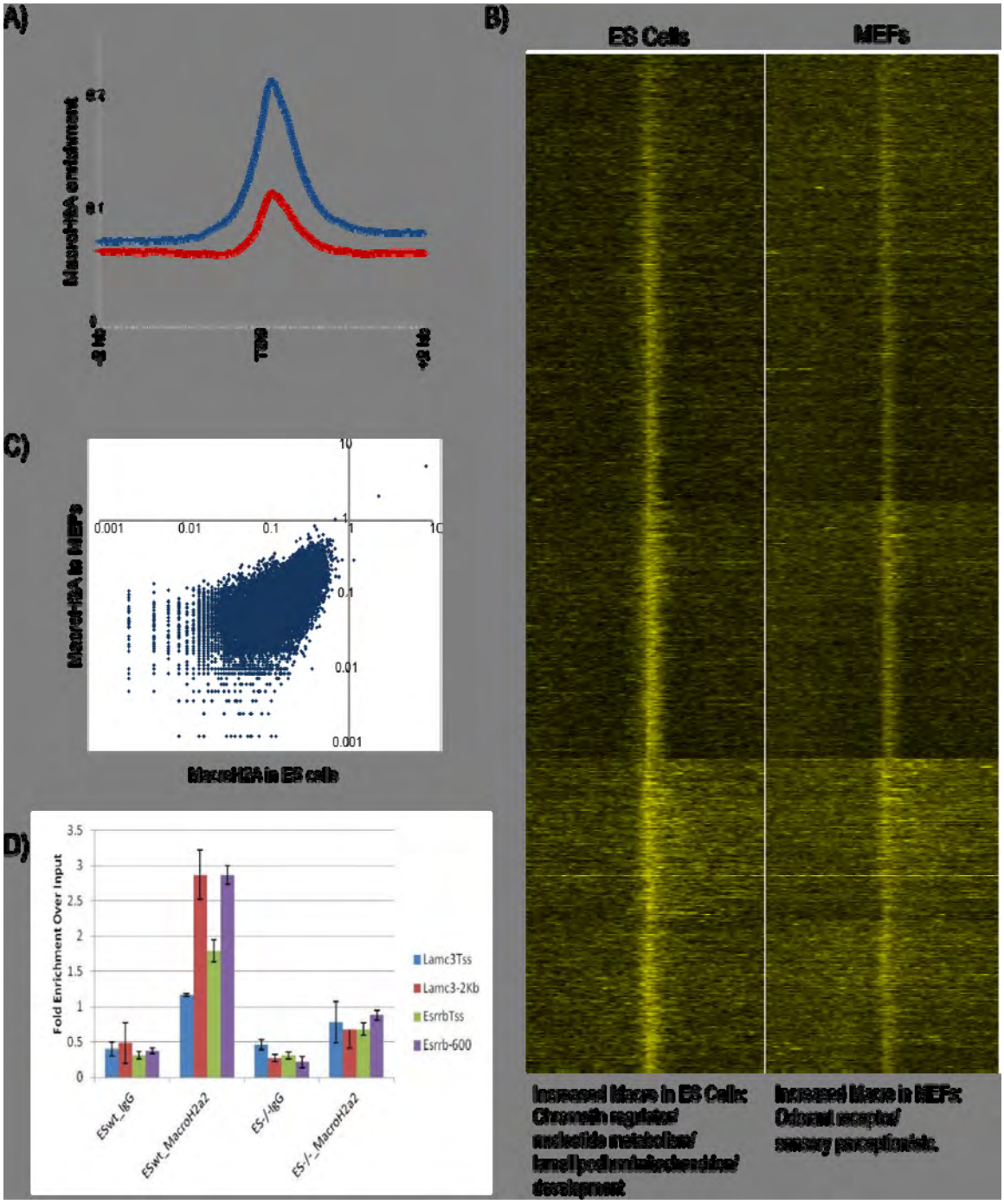

A)
B)
ES Cells
MEFs
C)
0.001
0.01
0.1
1
10
0.001
0.01
0.1
D)
Fold Enrichment Over Input
3.5
3
2.5
2
1.5
1
0.5
0
Lamc3Tss
Lamc3-2Kb
EsrrbTss
Esrrb-600
ESwt_IgG
ESwt_MacroH2a2
ES-/-IgG
ES-/-_MacroH2a2

Figure 3-5

Figure 3-5. Genome wide MacroH2A2 localization in mES cells and MEFs

A) Genome wide averaged localization of endogenous MacroH2A around the TSSs, in mouse ES cells and MEFs (their differentiated counter parts).

B) In both mES and MEFs endogenous MacroH2A shows similar overall localization profiles when aligned by the TSS. High enrichment of endogenous MAcroH2A around the TSS further support the integrity of high turnover profile observed in time course experiments (**Figure3-4A**) and the observation is not driven by overexpression. Endogenous MacroH2A is enriched at genes with different GO terms in ES and MEF cells, chromatin regulators, nucleotide metabolism, developmental genes in ES versus; odorant receptors and sensory perception in MEF cells.

C) Shows scatter plot of MacroH2A enrichment in MEF versus ES cells.

D) ChIP-qPCR validation of endogenous MacroH2A occupancy in wild type ES cells compared with macroH2A2 null ES cells, as negative control.

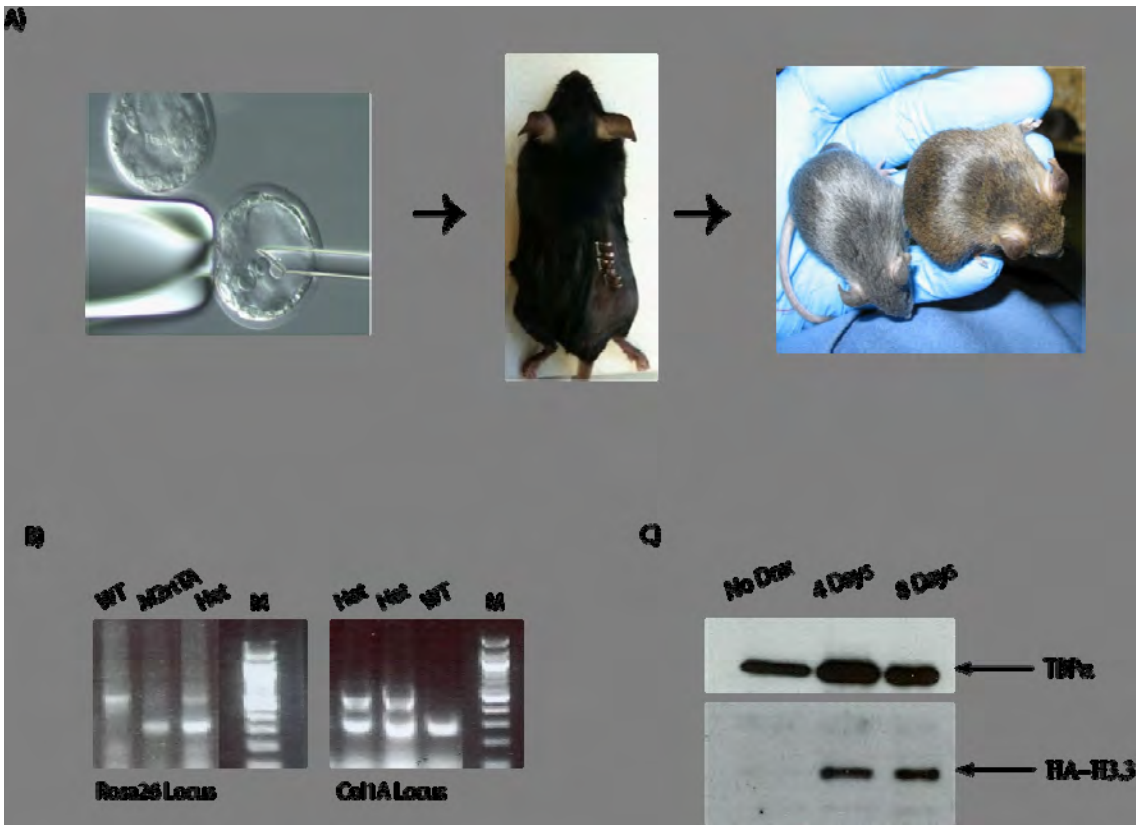

A)
B)
WT
Het
Het
M
Het
Het
WT
M
Rosa26 Locus
Col1A Locus
C)
No Dox
4 Days
8 Days
TRFs
HA-H3.3

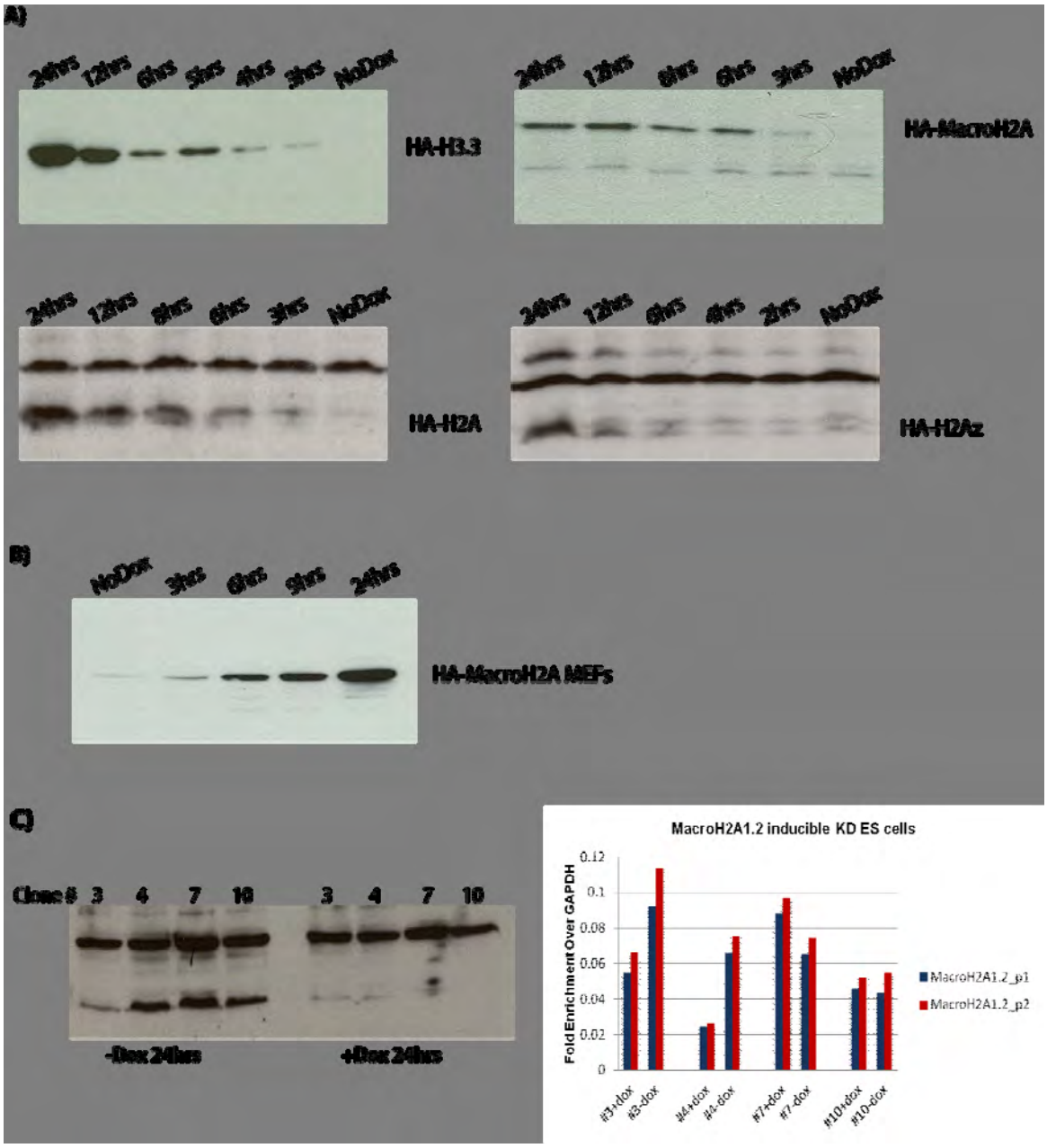
Figure 3-6

Figure 3-6. Generation of turnover mice.

A) HA-H3.3 ES cells were injected into Black6 X DBA F2 generation blastocysts (which have black coat color), and embryos were implanted into pseudopregnant females. We obtained multiple chimeric animals with high levels of contribution from the HA-H3.3 ES strain (which have agouti coat color) background, evident by enrichment of agouti coat color in F1 generation. (**Injection picture is from** the Transgenic/Gene Targeting Facility of Oregon Health & Science University, www.ohsu.edu/research/transgenics/images.shtml)

B) F2 generation genotyping showed that several of these animals had germline contribution from the tet-inducible ES cells. Wild type allele of Rosa26 locus where the M2rtTA resides has PCR product of ~500 base pairs (bp) while the M2rtTA containing allele has product size of ~300 bp. Wild type allele of Col1A locus has PCR product size of 331 bp, while the tet-inducible cassette integrated allele has product size of 551 bp.

C) Animals heterozygous for both Rosa26 and Col1A locus induced with feeding with doxycycline containing water (2mg/L doxycycline with 10g/L sucrose) for several days and animals sacrificed at different time points. Various tissues harvested, washed with PBS and flash frozen in liquid nitrogen. Expression of HA-H3.3 is observed in 4 days and 8 days induced animals but not in No Dox control animal, nuclear extracts from liver tissues. TBP-α is used as nuclear protein loading control.



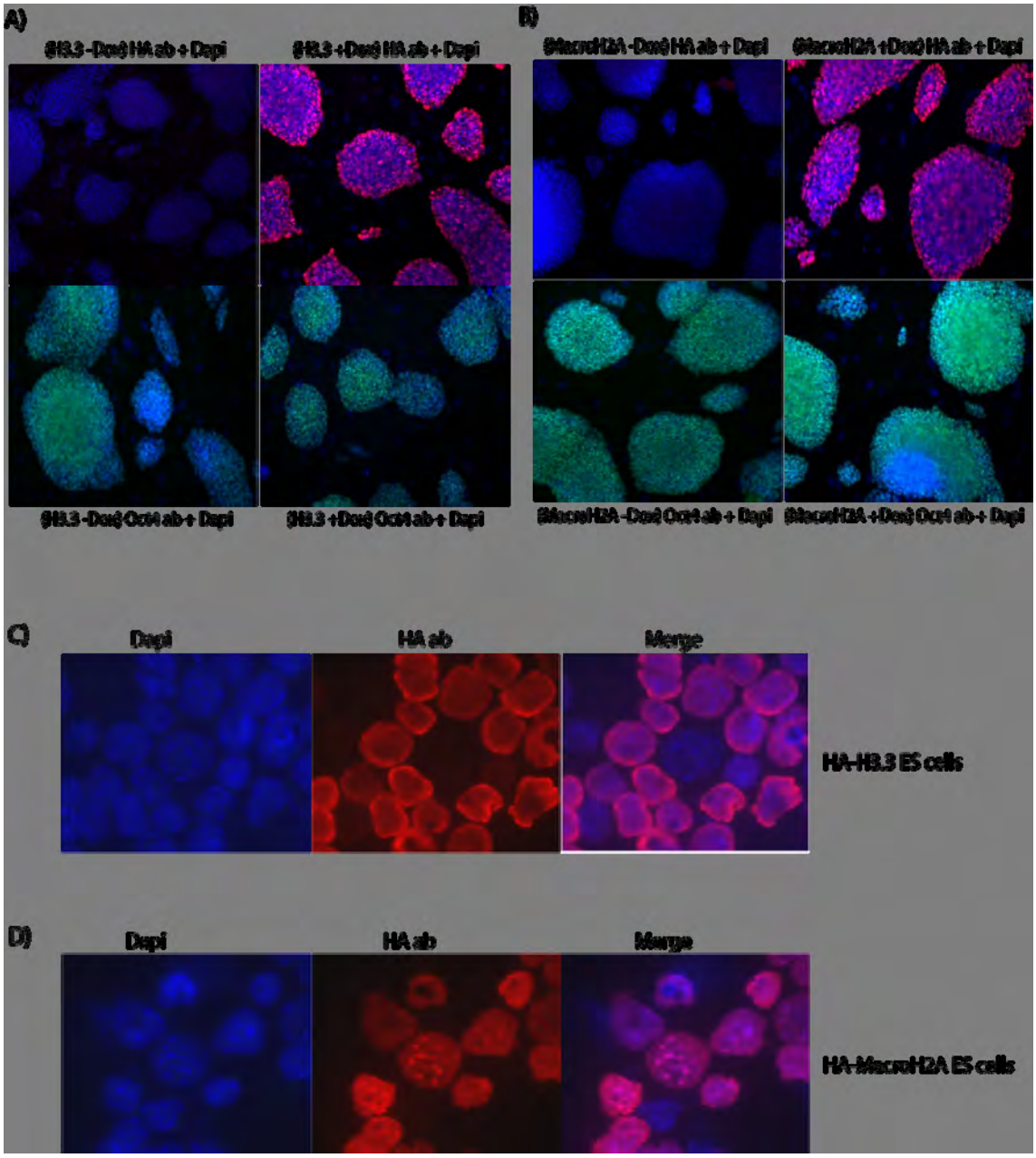
Supplemental Figure 3-1

Supplemental Figure 3-1. Confirmation of targeting construct integrity via transient transfection and other confirmed stable cell lines

A) Several constructs including HA tagged H2Az and H3.1 confirmed via transient transfection of rtTA expressing, Tet-ON PC12 cells. Those constructs were tried for targeting as well yet did not yield in genomically integrated stable ES cells (data not shown).

B) MEFs isolated from chimeras express HA tagged MacroH2A.

C) On the other hand we managed to generate stable inducible knockdown of macroH2A1.2; later which we had injected into blastocysts and obtained transgenic animals.



Supplemental Figure 3-2

Supplemental Figure 3-2. Ectopic expression of histone variants does not affect Oct4 levels and tagged variants can integrate into ES cell chromatin.

A) ES cell colonies, bearing inducible HA tagged H3.3, are immune fluorescent (IF) stained either with HA (red channel) or with Oct4 (green channel) antibody; samples counter stained with DAPI (blue channel). Both uninduced (-Dox) and 24hrs induced (+Dox) colonies showed similar Oct4 levels.

B) Similar results observed as in Supp Figure3-2A for HA tagged MacroH2A bearing ES cell colonies.

C) In order to have higher resolution images of the ES cells expressing tagged variants single cell suspensions were prepared and cytospinned on to positively charged glass slides. Cells IF stained with HA (red channel) and DAPI (blue channel). H3.3 enrichment show more peripheral enrichment in the nucleus

D) MacroH2A more significantly enriched in DAPI dense loci consistent with its heterochromatin association.

CHAPTER 4

Final Conclusions and General Discussion

The work presented in this dissertation characterizes the dynamic regulation of embryonic stem cell chromatin. We approached our studies by focusing on chromatin remodeling, DNA modifications and histone variants, their genome-wide localization and turnover. Our studies, described in chapter 2, revealed novel relationships between coactivator and corepressor complexes working antagonistically in the regulation of a common group of target genes in ES cells. Moreover, they revealed the interdependencies between chromatin remodeling, DNA methylation, DNA hydroxymethylation and the hydroxymethylating enzyme Tet1. On the other hand, our ES cell histone turnover system (chapter 3), provided deeper insight into both the histone turnover phenomenon and histone dynamics of different histone variants in pluripotent chromatin and in comparison with its differentiated counterparts.

Mbd3/NURD and BAF complexes antagonistically regulate common targets

Our results demonstrated that two integral components of two remodeling complexes; one mostly associated with repression and the other with activation, Mbd3 and Brg1, respectively, co-occupy several hundred genes. RNAi mediated depletion of these factors revealed that this co-occupancy is functional. While Mbd3 depletion resulted in increased expression; Brg1 depletion caused decreased expression of common target genes. Interestingly a double KD of these factors resulted in wild type-like expression profile, suggesting a role for this antagonistic relationship in fine tuning the expression levels of these genes. Our findings dovetail the theme of “bivalent” chromatin domains, in which

activation- and repression- associated posttranslational histone modifications that are produced by two antagonistic chromatin modifying complexes, H3K4me3 and H3K27me3, respectively, co-occupy developmentally important regulatory genes (Bernstein et al., 2006). Unlike H3K4me3 levels which appear to be dispensable for expression of most genes in ES cells (Jiang et al., 2011), Mbd3 and Brg1 exhibit direct opposing effects on RNA Polymerase II recruitment and gene expression at several hundred genes. Together, our data support the idea that chromatin regulatory factors with opposing functions maintain a transcriptionally permissive state of pluripotent chromatin while keeping expression levels of target genes in control.

Genome wide Mbd3 localization requires both Brg1 and Tet1 in ES cells

Despite its methyl binding domain (MBD), Mbd3 is deficient in binding to methylated DNA both *in vitro* and *in vivo* (Hendrich and Bird, 1998; Saito and Ishikawa, 2002) due to the substitution of a phenylalanine for a tyrosine in the MBD domain (Saito and Ishikawa, 2002). Intriguingly, this substitution is conserved between mouse and human Mbd3 methyl binding domains. Owing to this conservation and the phenotypic similarities observed upon depletion of Mbd3 and Tet1, the enzyme which hydrolyzes 5mC to 5hmC (Kaji et al., 2006); (Koh et al., 2011; Zhu et al., 2009) we hypothesized the potential role of hydroxymethylated DNA in Mbd3 recruitment.

Consistent with this hypothesis, our findings showed similar genome wide localization patterns of Mbd3 and Tet1 in ES cells. Both proteins are associated with CpG rich Polycomb bound promoters, while the highest Mbd3 enrichment

corresponds to genes with high levels of hydroxymethylation (Wu et al., 2011a). Moreover, Tet1 depletion causes genome wide Mbd3 de-localization, suggesting that hydroxymethylation plays a role in Mbd3 recruitment *in vivo*. Our *in vitro* findings demonstrated that Mbd3 preferentially binds to 5hmC-containing probes relative to 5mC-containing probes. Given the increase in 5mC observed in Tet1KD cells at loci that lose 5hmC (Wu et al., 2011a); (Ficz et al., 2011b) and our observation that 5mC inhibits Mbd3 binding, we speculate that genomic regions with 5mC intermixed with 5hmC might be unfavorable for Mbd3 binding. Future studies will dissect the details of how Mbd3/NURD differentially interacts with unmodified and hydroxymethylated DNA, and with hemi-modified or symmetrically modified DNA.

Our findings showed that genome wide localization of Mbd3 also requires Brg1 in mouse ES cells. It is conceivable, yet highly speculative that Brg1 may be functioning upstream of Tet1, facilitating its recruitment to target loci. This possibility is particularly intriguing, since our findings showed reduction of bulk levels of 5hmC in both Mbd3 and Brg1 depleted cells; that is comparable to Tet1 depletion *in vivo.* Further experiments are required to reveal the order of events that regulate Mbd3/NuRD recruitment with respect to hydroxymethylcytosine function.

Possible Mechanisms for the role of Mbd3 in the regulation of 5hmC levels

The reduced levels of hydroxymethylation in Mbd3 knockdowns suggest that Mbd3 plays an active role in regulating hydroxymethylation. Several possibilities exist that could explain our observations One possibility is that Mbd3

could bind to hydroxymethylated DNA and recruit Tet enzymes to hydroxylate adjacent methylcytosines. If this is the case Mbd3 depletion should cause an increase in 5mC levels similar to Tet1 depletion as it will also cause delocalization of Tet1. Another possibility is that Mbd3 could bind to hydroxymethylated loci and protect them from further steps in the demethylation pathway. If this is the case then Mbd3 depletion should cause an increase in unmodified cytosine levels while not affecting the 5mC levels. It is however clear that in either case, all three factors Brg1, Mbd3 and Tet1 “collaborate” to maintain hmC levels in ES cells. While Brg1 and Tet1 are necessary for Mbd3 localization, Mbd3, Brg1 and Tet1 are necessary for normal levels of cytosine hydroxymethylation. Although further studies are required to elucidate the detailed mechanistic basis for this interdependency, our findings describe Mbd3 as the first downstream effector that regulates expression of 5hmC-marked genes. Data presented in chapter 2 are consistent with a model in which Tet1-catalyzed hydroxymethylation serves to recruit the Mbd3/NURD complex, and thus Mbd3/NURD may be an effector that mediates some of the effects of hydroxymethylation on gene expression. Moreover our findings suggest that 5hmC plays a role in gene regulation beyond serving simply as an intermediate in a demethylation pathway. It will be of great interest in future studies to determine whether Mbd3 also plays a role in 5hmC biology in other contexts such as early embryos (Inoue and Zhang, 2011b; Iqbal et al., 2011), imprinting (Reese et al., 2007) or neurons (Kriaucionis and Heintz, 2009). Furthermore, dissecting the indispensible role of Mbd3 in embryonic development and

differentiation with respect to hydroxylmethylcytosine biology should provide us with deeper insights into epigenetic reprogramming.

H3.3 dynamics confirm conservation of histone turnover in mammals

Multiple factors and processes including, histone modifications, chromatin binding proteins, nucleosome positioning, assembly and remodeling of nucleosomes and replacement of canonical histones with histone variants contribute to the maintenance and regulation of active chromatin. Activation related nucleosome clearance at regulatory DNA elements is a common theme in chromatin regulation. A very well known example is facilitation of transcription factor binding after eviction of nucleosomes by the activity of ATP dependent chromatin remodelers at promoters of transcriptionally active genes. Another example is the correlation between DNaseI hypersensitivity sites and active regulatory DNA elements (Sabo et al., 2006). Recent studies implicated histone turnover as a mechanism to increase DNA accessibility. Genome wide studies in yeast and flies revealed unexpected histone replacement activity at regulatory regions, such as chromatin boundaries or Polycomb/Trithorax response elements (Deal et al., 2010; Dion et al., 2007; Mito et al., 2007). The coincidence of rapid histone turnover with specific DNA elements, most notably with promoter regions of actively transcribed genes, suggested a role for histone turnover in chromatin regulation, as higher histone turnover rates would provide a higher probability of DNA element exposure.

We therefore asked if the conservation of the histone turnover phenomenon extends to mammalian genomes. We chose to test this idea in

embryonic stem cells to gain more insight into the chromatin dynamics and potential role of histone turnover in pluripotent chromatin regulation. Also, comparing embryonic stem cell turnover profiles with various differentiated cell profiles would enable us to elucidate cell type-specific roles for histone turnover. We first focused on dynamics of the well-studied variant H3.3. H3.3 results served as proof of principle for our experimental system as our findings were consistent with previously published reports in terms of both genome wide localization in ES cells (Goldberg et al., 2010) and histone turnover profiles observed in flies (Deal et al., 2010; Mito et al., 2007). H3.3 turnover was most prominent at genes with high transcriptional activity and more rapid at the +1 nucleosomes than the -1 nucleosomes, a result that is consistent with the more rapid nucleosome replacement downstream of TSS measured by a metabolic labeling method in flies (Deal et al., 2010). Moreover our findings also showed higher turnover rates at Polycomb occupied genes which are usually targeted to CGI promoters in mammalian embryonic stem cells (Mikkelsen et al., 2007). Intriguingly, chromatin structure at CGI promoters were shown to be more accessible without a requirement for SWI/SNF nucleosome remodeling complexes (Ramirez-Carrozzi et al., 2009), suggesting a potential role for histone turnover in open chromatin formation at CGI promoters.

MacroH2A2 localization and dynamics in MEFs and ES cells

Our tagged MacroH2A2 turnover results revealed three distinct populations of MacroH2A containing chromatin that can be distinguished by their turnover rates. Two clusters with slow histone dynamics exhibit gradually

increasing enrichment of MacroH2A either upstream or downstream of TSS, respectively. Another cluster displays very rapid TSS proximal histone replacement while the third population shows no histone replacement activity even in later time points. Curiously, genes from all four clusters with different turnover rates are occupied by endogenous MacroH2A2 in wild type ES cells and at different levels in mouse embryonic fibroblasts. Our genome wide localization data showed that MacroH2A.2 widely localized to a large number of promoters in both pluripotent and differentiated cells, but that it is dynamic at only a subset of its binding sites with different turnover rates. These results suggest a potential involvement of different chaperone activities for distinct dynamic behaviors of MacroH2A2. We are currently investigating the functional consequences of the difference in MacroH2A2 dynamics by assaying gene expression defects in MacroH2A2 deficient ES cells. It will also be of great interest to see turnover rates in inducible MEFs as it will elucidate if there are differences in histone dynamic between different cell types with distinct chromatin landscapes. Moreover it will provide a better understanding of turnover dynamics of MacroH2A2 and their correlation with gene expression (ongoing experiments as well).

Inducible mice as a tool to study chromatin dynamics

Another exciting property of embryonic stem cells is that they are the cell type used to generate transgenic mice as they can contribute to the germline. So we took advantage of this developmental potential of embryonic stem cells to generate whole animals carrying the tetracycline-inducible epitope-tagged

histone variants. These animals have the potential to provide a unique and exciting resource for characterization of histone dynamics in different tissues *in vivo*. Moreover as doxycycline can cross the placenta and reach the fetus, these animals can be used to study different histone variant dynamics at different stages of early development *in vivo.*

ACKNOWLEDGEMENT

Chapter II (Mbd3/NURD complex & 5hmC Project):

Foremost, I am grateful to Dr. Thomas Fazzio, who carried out some of the most critical experiments for this project, including gene expression and gel shift experiments. His invaluable help especially in the preparation of most critical KD cell samples for my part of experiments allowed the projects progress successfully. I'd like to thank Dr. Eric Campeau for providing his protocol and expertise in virus preparation. I would like to acknowledge Dr. Christopher Lengner for teaching me stable ES cell generation and colony expansion for Mbd3 KD stable ES cells. I'd like to thank Chengjian Li for his assistance with TLC. I am also grateful to Dr. Chuan He for the generous gift of reagents, and to Dr. Chunxiao Song for his expertise and detailed protocol of 5hmC pull down.

Chapter III (Histone Turnover Project):

Foremost, I am deeply grateful to Dr. Christopher Lengner; all the stable cell lines and the inducible animals became a reality because of him, his expertise and his help. I am most grateful for his mentoring during stable cell line generation, MEF preparation, blastocyst injections and cytospin-IF experiments. I also would like to thank for his help during some of the time course experiments that I carried out in Jaenisch Lab. I would like to thank him and his graduate student Ryan Cedeno who derived the MEFs from inducible HA-tagged chimeras. I would like to extend

my gratitude to Jaenisch lab graduate student Michael Lodato who took care of our transgenic mice in Whitehead Institute, and set up some of the F1 and F2 breedings prior to their transfer to UMASSMED animal facility. I'd like thank Dr. Benjamin Carone for his help during the whole animal induction and tissue collection. I would also like to thank Dr. Mary Munson for her advice in Tris-Tricine gel that allowed me to compare ectopic expression levels of H3.3 to endogenous protein levels. I am very much grateful to Jui-Hung Hung who carried out deep sequencing data analysis. I'd like extend my sincere gratitude to Dr. Marta Radman-Livaja for helping me with further analyses of deep sequencing data, and also for her continuous input and discussions during data analyses writing of this project.

REFERENCES:

Abbott, D.W., Chadwick, B.P., Thambirajah, A.A., and Ausio, J. (2005). Beyond the Xi: macroH2A chromatin distribution and post-translational modification in an avian system. J Biol Chem *280*, 16437-16445.

Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol Cell *9*, 1191-1200.

Akiyama, T., Suzuki, O., Matsuda, J., and Aoki, F. (2011). Dynamic replacement of histone H3 variants reprograms epigenetic marks in early mouse embryos. PLoS Genet *7*, e1002279.

Andang, M., Hjerling-Leffler, J., Moliner, A., Lundgren, T.K., Castelo-Branco, G., Nanou, E., Pozas, E., Bryja, V., Halliez, S., Nishimaru, H., *et al.* (2008). Histone H2AX-dependent GABA(A) receptor regulation of stem cell proliferation. Nature *451*, 460-464.

Ang, Y.S., Tsai, S.Y., Lee, D.F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., *et al.* (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell *145*, 183-197.

Angelov, D., Bondarenko, V.A., Almagro, S., Menoni, H., Mongelard, F., Hans, F., Mietton, F., Studitsky, V.M., Hamiche, A., Dimitrov, S., *et al.* (2006). Nucleolin is a histone chaperone with FACT-like activity and assists remodeling of nucleosomes. Embo J *25*, 1669-1679.

Angelov, D., Molla, A., Perche, P.Y., Hans, F., Cote, J., Khochbin, S., Bouvet, P., and Dimitrov, S. (2003). The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling. Mol Cell *11*, 1033-1041.

Arents, G., and Moudrianakis, E.N. (1995). The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. Proc Natl Acad Sci U S A *92*, 11170-11174.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Barski, A., Jothi, R., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., and Zhao, K. (2009). Chromatin poises miRNA- and protein-coding genes for expression. Genome Res *19*, 1742-1751.

Baylin, S.B., and Jones, P.A. (2011). A decade of exploring the cancer epigenome - biological and translational implications. Nature reviews *11*, 726-734.

Beard, C., Hochedlinger, K., Plath, K., Wutz, A., and Jaenisch, R. (2006). Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. Genesis *44*, 23-28.

Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., and Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. Science *301*, 1090-1093.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bhutani, N., Burns, D.M., and Blau, H.M. (2011). DNA demethylation dynamics. Cell *146*, 866-872.

Bogdanovic, O., and Veenstra, G.J. (2009). DNA methylation and methyl-CpG binding proteins: developmental requirements and function. Chromosoma *118*, 549-565.
Bonnefoy, E., Orsi, G., Couble, P., and Loppin, B. (2005). The essential role of Drosophila HIRA for de novo assembly of paternal chromatin at fertilization. PLoS Genetics *preprint*, e182.
Boyer, L.A., Latek, R.R., and Peterson, C.L. (2004). The SANT domain: a unique histone-tail-binding module? Nat Rev Mol Cell Biol *5*, 158-163.
Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., *et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.
Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., *et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.
Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev *20*, 1123-1136.
Bramlage, B., Kosciessa, U., and Doenecke, D. (1997). Differential expression of the murine histone genes H3.3A and H3.3B. Differentiation; research in biological diversity *62*, 13-20.
Braunschweig, U., Hogan, G.J., Pagie, L., and van Steensel, B. (2009). Histone H1 binding is inhibited by histone variant H3.3. Embo J *28*, 3635-3645.
Bultman, S.J., Gebuhr, T.C., Pan, H., Svoboda, P., Schultz, R.M., and Magnuson, T. (2006). Maternal BRG1 regulates zygotic genome activation in the mouse. Genes Dev *20*, 1744-1754.
Buschbeck, M., Uribesalgo, I., Wibowo, I., Rue, P., Martin, D., Gutierrez, A., Morey, L., Guigo, R., Lopez-Schier, H., and Di Croce, L. (2009). The histone variant macroH2A is an epigenetic regulator of key developmental genes. Nat Struct Mol Biol *16*, 1074-1079.
Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P., *et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.
Celeste, A., Petersen, S., Romanienko, P.J., Fernandez-Capetillo, O., Chen, H.T., Sedelnikova, O.A., Reina-San-Martin, B., Coppola, V., Meffre, E., Difilippantonio, M.J., *et al.* (2002). Genomic instability in mice lacking histone H2AX. Science *296*, 922-927.
Chaboute, M.E., Chaubet, N., Gigot, C., and Philipps, G. (1993). Histones and histone genes in higher plants: structure and genomic organization. Biochimie *75*, 523-531.
Chadwick, B.P., and Willard, H.F. (2001a). Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. Hum Mol Genet *10*, 1101-1113.
Chadwick, B.P., and Willard, H.F. (2001b). A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. J Cell Biol *152*, 375-384.
Chadwick, B.P., and Willard, H.F. (2002). Cell cycle-dependent localization of macroH2A in chromatin of the inactive X chromosome. J Cell Biol *157*, 1113-1123.

Chakravarthy, S., Gundimella, S.K., Caron, C., Perche, P.Y., Pehrson, J.R., Khochbin, S., and Luger, K. (2005a). Structural characterization of the histone variant macroH2A. Mol Cell Biol *25*, 7616-7624.
Chakravarthy, S., Park, Y.J., Chodaparambil, J., Edayathumangalam, R.S., and Luger, K. (2005b). Structure and dynamic properties of nucleosome core particles. FEBS Lett *579*, 895-898.
Cho, R.J., Campbell, M.J., Winzeler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J., *et al.* (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell *2*, 65-73.
Cho, R.J., Huang, M., Campbell, M.J., Dong, H., Steinmetz, L., Sapinoso, L., Hampton, G., Elledge, S.J., Davis, R.W., and Lockhart, D.J. (2001). Transcriptional regulation and function during the human cell cycle. Nat Genet *27*, 48-54.
Chowdhary, R., Ali, R.A., Albig, W., Doenecke, D., and Bajic, V.B. (2005). Promoter modeling: the case study of mammalian histone promoters. Bioinformatics *21*, 2623-2628.
Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G., and Li, B.F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science *277*, 1996-2000.
Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu Rev Biochem *78*, 273-304.
Clarkson, M.J., Wells, J.R., Gibson, F., Saint, R., and Tremethick, D.J. (1999). Regions of variant histone His2AvD required for Drosophila development. Nature *399*, 694-697.
Collins, K.A., Castillo, A.R., Tatsutani, S.Y., and Biggins, S. (2005). De novo kinetochore assembly requires the centromeric histone H3 variant. Mol Biol Cell *16*, 5649-5660.
Collins, K.A., Furuyama, S., and Biggins, S. (2004). Proteolysis contributes to the exclusive centromere localization of the yeast Cse4/CENP-A histone H3 variant. Curr Biol *14*, 1968-1972.
Core, L.J., and Lis, J.T. (2008). Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science *319*, 1791-1792.
Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.
Cosma, M.P., Tanaka, T., and Nasmyth, K. (1999). Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. Cell *97*, 299-311.
Costanzi, C., and Pehrson, J.R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*, 599-601.
Costanzi, C., and Pehrson, J.R. (2001). MACROH2A2, a new member of the MARCOH2A core histone family. J Biol Chem *276*, 21776-21784.
Cote, J., Quinn, J., Workman, J.L., and Peterson, C.L. (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. Science *265*, 53-60.
Couldrey, C., Carlton, M.B., Nolan, P.M., Colledge, W.H., and Evans, M.J. (1999). A retroviral gene trap insertion into the histone 3.3A gene causes partial neonatal lethality,

stunted growth, neuromuscular deficits and male sub-fertility in transgenic mice. Hum Mol Genet *8*, 2489-2495.
Creppe, C., Janich, P., Cantarino, N., Noguera, M., Valero, V., Musulen, E., Douet, J., Posavec, M., Martin Caballero, J., Sumoy, L., *et al.* (2012). MacroH2A1 regulates the balance between self-renewal and differentiation commitment in embryonic and adult stem cells. Mol Cell Biol.
Creyghton, M.P., Markoulaki, S., Levine, S.S., Hanna, J., Lodato, M.A., Sha, K., Young, R.A., Jaenisch, R., and Boyer, L.A. (2008). H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. Cell *135*, 649-661.
Csankovszki, G., Nagy, A., and Jaenisch, R. (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. J Cell Biol *153*, 773-784.
Csankovszki, G., Panning, B., Bates, B., Pehrson, J.R., and Jaenisch, R. (1999). Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. Nat Genet *22*, 323-324.
Culhane, A.C., Thioulouse, J., Perriere, G., and Higgins, D.G. (2005). MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics *21*, 2789-2790.
Dawlaty, M.M., Ganz, K., Powell, B.E., Hu, Y.C., Markoulaki, S., Cheng, A.W., Gao, Q., Kim, J., Choi, S.W., Page, D.C., *et al.* (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell *9*, 166-175.
De Koning, L., Corpet, A., Haber, J.E., and Almouzni, G. (2007). Histone chaperones: an escort network regulating histone traffic. Nat Struct Mol Biol *14*, 997-1007.
Deal, R.B., Henikoff, J.G., and Henikoff, S. (2010). Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. Science *328*, 1161-1164.
Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev *25*, 1010-1022.
Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol *4*, P3.
Dion, M.F., Altschuler, S.J., Wu, L.F., and Rando, O.J. (2005). Genomic characterization reveals a simple histone H4 acetylation code. Proc Natl Acad Sci U S A.
Dion, M.F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N., and Rando, O.J. (2007). Dynamics of replication-independent histone turnover in budding yeast. Science *315*, 1405-1408.
Doyen, C.M., An, W., Angelov, D., Bondarenko, V., Mietton, F., Studitsky, V.M., Hamiche, A., Roeder, R.G., Bouvet, P., and Dimitrov, S. (2006a). Mechanism of polymerase II transcription repression by the histone variant macroH2A. Mol Cell Biol *26*, 1156-1164.
Doyen, C.M., Montel, F., Gautier, T., Menoni, H., Claudet, C., Delacour-Larose, M., Angelov, D., Hamiche, A., Bednar, J., Faivre-Moskalenko, C., *et al.* (2006b). Dissection of the unusual structural and functional properties of the variant H2A.Bbd nucleosome. Embo J *25*, 4234-4244.

Drane, P., Ouararhni, K., Depaux, A., Shuaib, M., and Hamiche, A. (2010). The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. Genes Dev *24*, 1253-1265.
Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H., *et al.* (2008). Global transcription in pluripotent embryonic stem cells. Cell Stem Cell *2*, 437-447.
Faast, R., Thonglairoam, V., Schulz, T.C., Beall, J., Wells, J.R., Taylor, H., Matthaei, K., Rathjen, P.D., Tremethick, D.J., and Lyons, I. (2001). Histone variant H2A.Z is required for early mammalian development. Curr Biol *11*, 1183-1187.
Fan, J.Y., Gordon, F., Luger, K., Hansen, J.C., and Tremethick, D.J. (2002). The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states. Nat Struct Biol *9*, 172-176.
Fan, J.Y., Rangasamy, D., Luger, K., and Tremethick, D.J. (2004). H2A.Z alters the nucleosome surface to promote HP1alpha-mediated chromatin fiber folding. Mol Cell *16*, 655-661.
Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L., and Skoultchi, A.I. (2005). Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. Cell *123*, 1199-1212.
Fazzio, T.G., Huff, J.T., and Panning, B. (2008a). Chromatin regulation Tip(60)s the balance in embryonic stem cell self-renewal. Cell Cycle *7*, 3302-3306.
Fazzio, T.G., Huff, J.T., and Panning, B. (2008b). An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. Cell *134*, 162-174.
Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., *et al.* (2010). Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A *107*, 8689-8694.
Fernandez-Capetillo, O., Lee, A., Nussenzweig, M., and Nussenzweig, A. (2004). H2AX: the histone guardian of the genome. DNA repair *3*, 959-967.
Fernandez-Capetillo, O., Mahadevaiah, S.K., Celeste, A., Romanienko, P.J., Camerini-Otero, R.D., Bonner, W.M., Manova, K., Burgoyne, P., and Nussenzweig, A. (2003). H2AX is required for chromatin remodeling and inactivation of sex chromosomes in male mouse meiosis. Dev Cell *4*, 497-508.
Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011a). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature.
Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011b). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature *473*, 398-402.
Flanagan, J.F., Mi, L.Z., Chruszcz, M., Cymborowski, M., Clines, K.L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. (2005). Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature *438*, 1181-1185.
Flaus, A., and Owen-Hughes, T. (2004). Mechanisms for ATP-dependent chromatin remodelling: farewell to the tuna-can octamer? Curr Opin Genet Dev *14*, 165-173.

Francis, N.J., Follmer, N.E., Simon, M.D., Aghia, G., and Butler, J.D. (2009). Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. Cell *137*, 110-122.
Frank, D., Doenecke, D., and Albig, W. (2003). Differential expression of human replacement and cell cycle dependent H3 histone genes. Gene *312*, 135-143.
Friedman, J.H. (1991). Multivariate Adaptive Regression Splines. Ann Statist *19*, 1-97.
Gamble, M.J., Frizzell, K.M., Yang, C., Krishnakumar, R., and Kraus, W.L. (2010). The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing. Genes Dev *24*, 21-32.
Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M.J., Heidersbach, A., Ramalho-Santos, J., McManus, M.T., Plath, K., Meshorer, E., *et al.* (2009). Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature *460*, 863-868.
Gautier, T., Abbott, D.W., Molla, A., Verdel, A., Ausio, J., and Dimitrov, S. (2004). Histone variant H2ABbd confers lower stability to the nucleosome. EMBO reports *5*, 715-720.
Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., *et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol *5*, R80.
Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., *et al.* (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell *140*, 678-691.
Goldmark, J.P., Fazzio, T.G., Estep, P.W., Church, G.M., and Tsukiyama, T. (2000). The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p. Cell *103*, 423-433.
Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.
Guertin, M.J., Petesch, S.J., Zobeck, K.L., Min, I.M., and Lis, J.T. (2010). Drosophila heat shock system as a general model to investigate transcriptional regulation. Cold Spring Harb Symp Quant Biol *75*, 1-9.
Gunjan, A., Paik, J., and Verreault, A. (2006). The emergence of regulated histone proteolysis. Curr Opin Genet Dev *16*, 112-118.
Gunjan, A., and Verreault, A. (2003). A Rad53 kinase-dependent surveillance mechanism that regulates histone protein levels in S. cerevisiae. Cell *115*, 537-549.
Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U.C., Cesari, F., Lee, C., Almouzni, G., Schneider, R., and Surani, M.A. (2008). Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature *452*, 877-881.
Hake, S.B., Garcia, B.A., Duncan, E.M., Kauer, M., Dellaire, G., Shabanowitz, J., Bazett-Jones, D.P., Allis, C.D., and Hunt, D.F. (2006). Expression patterns and post-translational modifications associated with mammalian histone H3 variants. J Biol Chem *281*, 559-568.
Hardy, S., Jacques, P.E., Gevry, N., Forest, A., Fortin, M.E., Laflamme, L., Gaudreau, L., and Robert, F. (2009). The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A.Z deposition. PLoS Genet *5*, e1000687.

Hassan, A.H., Prochasson, P., Neely, K.E., Galasinski, S.C., Chandy, M., Carrozza, M.J., and Workman, J.L. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell *111*, 369-379.
Hata, K., Okano, M., Lei, H., and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development *129*, 1983-1993.
Hatch, C.L., and Bonner, W.M. (1988). Sequence of cDNAs for mammalian H2A.Z, an evolutionarily diverged but highly conserved basal histone H2A isoprotein species. Nucleic Acids Res *16*, 1113-1124.
Heintz, N., Sive, H.L., and Roeder, R.G. (1983). Regulation of human histone gene expression: kinetics of accumulation and changes in the rate of synthesis and in the half-lives of individual histone mRNAs during the HeLa cell cycle. Mol Cell Biol *3*, 539-550.
Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., *et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.
Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol *18*, 6538-6547.
Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. Nat Rev Genet *9*, 15-26.
Henikoff, S., and Ahmad, K. (2005). Assembly of variant histones into chromatin. Annu Rev Cell Dev Biol *21*, 133-153.
Henikoff, S., Furuyama, T., and Ahmad, K. (2004). Histone variants, nucleosome assembly and epigenetic inheritance. Trends Genet *20*, 320-326.
Hentschel, C.C., and Birnstiel, M.L. (1981). The organization and expression of histone gene families. Cell *25*, 301-313.
Heo, K., Kim, H., Choi, S.H., Choi, J., Kim, K., Gu, J., Lieber, M.R., Yang, A.S., and An, W. (2008). FACT-mediated exchange of histone variant H2AX regulated by phosphorylation of H2AX and ADP-ribosylation of Spt16. Mol Cell *30*, 86-97.
Ho, L., Jothi, R., Ronan, J.L., Cui, K., Zhao, K., and Crabtree, G.R. (2009a). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc Natl Acad Sci U S A *106*, 5187-5191.
Ho, L., Ronan, J.L., Wu, J., Staahl, B.T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A.I., Ranish, J., and Crabtree, G.R. (2009b). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. Proc Natl Acad Sci U S A *106*, 5181-5186.
Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics. Scand J Stat *6*, 65-70.
Howman, E.V., Fowler, K.J., Newson, A.J., Redward, S., MacDonald, A.C., Kalitsis, P., and Choo, K.H. (2000). Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. Proc Natl Acad Sci U S A *97*, 1148-1153.
Hu, Q., Friedrich, A.M., Johnson, L.V., and Clegg, D.O. (2010). Memory in induced pluripotent stem cells: reprogrammed human retinal-pigmented epithelial cells show tendency for spontaneous redifferentiation. Stem cells (Dayton, Ohio) *28*, 1981-1991.
Hu, Y., Chopra, V., Chopra, R., Locascio, J.J., Liao, Z., Ding, H., Zheng, B., Matson, W.R., Ferrante, R.J., Rosas, H.D., *et al.* (2011). Transcriptional modulator H2A histone

family, member Y (H2AFY) marks Huntington disease activity in man and mouse. Proc Natl Acad Sci U S A *108*, 17141-17146.
Illingworth, R., Kerr, A., Desousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., *et al.* (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol *6*, e22.
Inoue, A., and Zhang, Y. (2011a). Replication-Dependent Loss of 5-Hydroxymethylcytosine in Mouse Preimplantation Embryos. Science.
Inoue, A., and Zhang, Y. (2011b). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science *334*, 194.
Iqbal, K., Jin, S.G., Pfeifer, G.P., and Szabo, P.E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci U S A *108*, 3642-3647.
Ishibashi, T., Li, A., Eirin-Lopez, J.M., Zhao, M., Missiaen, K., Abbott, D.W., Meistrich, M., Hendzel, M.J., and Ausio, J. (2010). H2A.Bbd: an X-chromosome-encoded histone involved in mammalian spermiogenesis. Nucleic Acids Res *38*, 1780-1789.
Isogai, Y., Keles, S., Prestel, M., Hochheimer, A., and Tjian, R. (2007). Transcription of histone gene cluster by differential core-promoter factors. Genes Dev *21*, 2936-2949.
Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129-1133.
Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.
Jiang, H., Shukla, A., Wang, X., Chen, W.Y., Bernstein, B.E., and Roeder, R.G. (2011). Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. Cell *144*, 513-525.
Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet *41*, 941-945.
Jin, S.G., Kadam, S., and Pfeifer, G.P. (2010). Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res *38*, e125.
Kaji, K., Caballero, I.M., MacLeod, R., Nichols, J., Wilson, V.A., and Hendrich, B. (2006). The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nat Cell Biol *8*, 285-292.
Kapoor, A., Goldberg, M.S., Cumberland, L.K., Ratnakumar, K., Segura, M.F., Emanuel, P.O., Menendez, S., Vardabasso, C., Leroy, G., Vidal, C.I., *et al.* (2010). The histone variant macroH2A suppresses melanoma progression through regulation of CDK8. Nature *468*, 1105-1109.
Kim, J., Sif, S., Jones, B., Jackson, A., Koipally, J., Heller, E., Winandy, S., Viel, A., Sawyer, A., Ikeda, T., *et al.* (1999). Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. Immunity *10*, 345-355.
Kim, J., Woo, A.J., Chu, J., Snow, J.W., Fujiwara, Y., Kim, C.G., Cantor, A.B., and Orkin, S.H. (2010a). A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell *143*, 313-324.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I., *et al.* (2010b). Epigenetic memory in induced pluripotent stem cells. Nature *467*, 285-290.
Kim, K., Zhao, R., Doi, A., Ng, K., Unternaehrer, J., Cahan, P., Huo, H., Loh, Y.H., Aryee, M.J., Lensch, M.W., *et al.* (2011). Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. Nat Biotechnol *29*, 1117-1119.
Kimura, H., and Cook, P.R. (2001). Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. J Cell Biol *153*, 1341-1353.
Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci *31*, 89-97.
Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C.A., Mostoslavsky, G., *et al.* (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell *8*, 200-213.
Konev, A.Y., Tribus, M., Park, S.Y., Podhraski, V., Lim, C.Y., Emelyanov, A.V., Vershilova, E., Pirrotta, V., Kadonaga, J.T., Lusser, A., *et al.* (2007). CHD1 motor protein is required for deposition of histone variant H3.3 into chromatin in vivo. Science *317*, 1087-1090.
Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.
Kornberg, R.D., and Thomas, J.O. (1974). Chromatin structure; oligomers of the histones. Science *184*, 865-868.
Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.
Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929-930.
Kustatscher, G., Hothorn, M., Pugieux, C., Scheffzek, K., and Ladurner, A.G. (2005). Splicing regulates NAD metabolite binding to histone macroH2A. Nat Struct Mol Biol *12*, 624-625.
Le Guezennec, X., Vermeulen, M., Brinkman, A.B., Hoeijmakers, W.A., Cohen, A., Lasonder, E., and Stunnenberg, H.G. (2006). MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. Mol Cell Biol *26*, 843-851.
Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., *et al.* (2006a). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.
Lee, T.I., Johnstone, S.E., and Young, R.A. (2006b). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc *1*, 729-748.
Lessard, J.A., and Crabtree, G.R. (2010). Chromatin regulatory mechanisms in pluripotency. Annu Rev Cell Dev Biol *26*, 503-532.
Lewis, P.W., Elsaesser, S.J., Noh, K.M., Stadler, S.C., and Allis, C.D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. Proc Natl Acad Sci U S A *107*, 14075-14080.

Li, Q., Zhou, H., Wurtele, H., Davies, B., Horazdovsky, B., Verreault, A., and Zhang, Z. (2008). Acetylation of histone H3 lysine 56 regulates replication-coupled nucleosome assembly. Cell *134*, 244-255.
Liu, C.L., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S.L., Friedman, N., and Rando, O.J. (2005). Single-Nucleosome Mapping of Histone Modifications in S. cerevisiae. PLoS Biol *3*, e328.
Loebel, D.A.F., Watson, C.M., De Young, R.A., and Tam, P.P.L. (2003). Lineage choice and differentiation in mouse embryos and embryonic stem cells. Developmental biology *264*, 1-14.
Loppin, B., Bonnefoy, E., Anselme, C., Laurencon, A., Karr, T.L., and Couble, P. (2005). The histone H3.3 chaperone HIRA is essential for chromatin assembly in the male pronucleus. Nature *437*, 1386-1390.
Lorch, Y., Maier-Davis, B., and Kornberg, R.D. (2006). Chromatin remodeling by nucleosome disassembly in vitro. Proc Natl Acad Sci U S A *103*, 3090-3093.
Loyola, A., Bonaldi, T., Roche, D., Imhof, A., and Almouzni, G. (2006). PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. Mol Cell *24*, 309-316.
Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997a). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.
Luger, K., Rechsteiner, T.J., Flaus, A.J., Waye, M.M., and Richmond, T.J. (1997b). Characterization of nucleosome core particles containing histone proteins made in bacteria. J Mol Biol *272*, 301-311.
Luk, E., Ranjan, A., Fitzgerald, P.C., Mizuguchi, G., Huang, Y., Wei, D., and Wu, C. (2010). Stepwise histone replacement by SWR1 requires dual activation with histone H2A.Z and canonical nucleosome. Cell *143*, 725-736.
Mahajan, M.C., Narlikar, G.J., Boyapaty, G., Kingston, R.E., and Weissman, S.M. (2005). Heterogeneous nuclear ribonucleoprotein C1/C2, MeCP1, and SWI/SNF form a chromatin remodeling complex at the beta-globin locus control region. Proc Natl Acad Sci U S A *102*, 15012-15017.
Malik, H.S., and Henikoff, S. (2003). Phylogenomics of the nucleosome. Nat Struct Biol *10*, 882-891.
Marino-Ramirez, L., Jordan, I.K., and Landsman, D. (2006). Multiple independent evolutionary solutions to core histone gene regulation. Genome Biol *7*, R122.
Marino-Ramirez, L., Kann, M.G., Shoemaker, B.A., and Landsman, D. (2005). Histone structure and nucleosome stability. Expert review of proteomics *2*, 719-729.
Martens, J.A., and Winston, F. (2002). Evidence that Swi/Snf directly represses transcription in S. cerevisiae. Genes Dev *16*, 2231-2236.
Marzluff, W.F., Gongidi, P., Woods, K.R., Jin, J., and Maltais, L.J. (2002). The human and mouse replication-dependent histone genes. Genomics *80*, 487-498.
Marzluff, W.F., Wagner, E.J., and Duronio, R.J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet *9*, 843-854.
Maxson, R., Cohn, R., Kedes, L., & Mohun, T. (1983). Maxson, R., Cohn, R., Kedes, L., & Mohun, T. (1983). Maxson, R., Cohn, R., Kedes, L., and Mohun, T. (1983). Expression and organization of histone genes. Annu Rev Genet *17*, 239-277.

McDonel, P., Costello, I., and Hendrich, B. (2009). Keeping things quiet: roles of NuRD and Sin3 co-repressor complexes during mammalian development. The international journal of biochemistry & cell biology *41*, 108-116.
McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc Natl Acad Sci U S A *101*, 1525-1530.
Meissner, A., Eminli, S., and Jaenisch, R. (2009). Derivation and manipulation of murine embryonic stem cells. Methods Mol Biol *482*, 3-19.
Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., *et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.
Meneghini, M.D., Wu, M., and Madhani, H.D. (2003). Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. Cell *112*, 725-736.
Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat Rev Mol Cell Biol *7*, 540-546.
Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Dev Cell *10*, 105-116.
Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49-55.
Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.
Mito, Y., Henikoff, J.G., and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. Nat Genet *37*, 1090-1097.
Mito, Y., Henikoff, J.G., and Henikoff, S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. Science *315*, 1408-1411.
Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell *30*, 755-766.
Morillon, A., Karabetsou, N., O'Sullivan, J., Kent, N., Proudfoot, N., and Mellor, J. (2003). Isw1 chromatin remodeling ATPase coordinates transcription elongation and termination by RNA polymerase II. Cell *115*, 425-435.
Narlikar, G.J., Fan, H.Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475-487.
Nishikawa, S., Jakt, L.M., and Era, T. (2007). Embryonic stem-cell culture as a tool for developmental cell biology. Nat Rev Mol Cell Biol *8*, 502-507.
Nusinow, D.A., Hernandez-Munoz, I., Fazzio, T.G., Shah, G.M., Kraus, W.L., and Panning, B. (2007). Poly(ADP-ribose) polymerase 1 is inhibited by a histone H2A variant, MacroH2A, and contributes to silencing of the inactive X chromosome. J Biol Chem *282*, 12851-12859.
Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M., and Shirakawa, M. (2001). Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. Cell *105*, 487-497.

Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). Science *183*, 330-332.
Onder, T.T., Kara, N., Cherry, A., Sinha, A.U., Zhu, N., Bernt, K.M., Cahan, P., Marcarci, B.O., Unternaehrer, J., Gupta, P.B., *et al.* (2012). Chromatin-modifying enzymes as modulators of reprogramming. Nature *483*, 598-602.
Orlando, V. (2003). Polycomb, epigenomes, and control of cell identity. Cell *112*, 599-606.
Orsi, G.A., Couble, P., and Loppin, B. (2009). Epigenetic and replacement roles of histone variant H3.3 in reproduction and development. The International journal of developmental biology *53*, 231-243.
Osley, M.A. (1991). The regulation of histone synthesis in the cell cycle. Annu Rev Biochem *60*, 827-861.
Papamichos-Chronakis, M., Watanabe, S., Rando, O.J., and Peterson, C.L. (2011). Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. Cell *144*, 200-213.
Park, Y.J., Dyer, P.N., Tremethick, D.J., and Luger, K. (2004). A new fluorescence resonance energy transfer approach demonstrates that the histone variant H2AZ stabilizes the histone octamer within the nucleosome. J Biol Chem *279*, 24274-24282.
Pehrson, J.R., Costanzi, C., and Dharia, C. (1997). Developmental and tissue expression patterns of histone macroH2A1 subtypes. J Cell Biochem *65*, 107-113.
Pehrson, J.R., and Fried, V.A. (1992). MacroH2A, a core histone containing a large nonhistone region. Science *257*, 1398-1400.
Peterson, C.L., and Workman, J.L. (2000). Promoter targeting and chromatin remodeling by the SWI/SNF complex. Curr Opin Genet Dev *10*, 187-192.
Pina, B., and Suau, P. (1987). Changes in histones H2A and H3 variant composition in differentiating and mature rat brain cortical neurons. Developmental biology *123*, 51-58.
Raisner, R.M., Hartley, P.D., Meneghini, M.D., Bao, M.Z., Liu, C.L., Schreiber, S.L., Rando, O.J., and Madhani, H.D. (2005). Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. Cell *123*, 233-248.
Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell *138*, 114-128.
Rando, O.J. (2007). Global patterns of histone modifications. Curr Opin Genet Dev *17*, 94-99.
Rando, O.J., and Ahmad, K. (2007). Rules and regulation in the primary structure of chromatin. Curr Opin Cell Biol *19*, 250-256.
Rando, O.J., and Chang, H.Y. (2009). Genome-wide views of chromatin structure. Annu Rev Biochem *78*, 245-271.
Ratnakumar, K., Duarte, L.F., Leroy, G., Hasson, D., Smeets, D., Vardabasso, C., Bonisch, C., Zeng, T., Xiang, B., Zhang, D.Y., *et al.* (2012). ATRX-mediated chromatin association of histone variant macroH2A1 regulates alpha-globin expression. Genes Dev *26*, 433-438.
Reese, K.J., Lin, S., Verona, R.I., Schultz, R.M., and Bartolomei, M.S. (2007). Maintenance of paternal methylation and repression of the imprinted H19 gene requires MBD3. PLoS Genet *3*, e137.

Ringrose, L., and Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development *134*, 223-232.
Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem *273*, 5858-5868.
Rufiange, A., Jacques, P.E., Bhat, W., Robert, F., and Nourani, A. (2007). Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. Mol Cell *27*, 393-405.
Sabo, P.J., Kuehn, M.S., Thurman, R., Johnson, B.E., Johnson, E.M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., *et al.* (2006). Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. Nat Methods *3*, 511-518.
Saha, A., Wittmeyer, J., and Cairns, B.R. (2002). Chromatin remodeling by RSC involves ATP-dependent DNA translocation. Genes Dev *16*, 2120-2134.
Saito, M., and Ishikawa, F. (2002). The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. J Biol Chem *277*, 35434-35439.
Sakai, A., Schwartz, B.E., Goldstein, S., and Ahmad, K. (2009). Transcriptional and developmental functions of the H3.3 histone variant in Drosophila. Curr Biol *19*, 1816-1820.
Sawatsubashi, S., Murata, T., Lim, J., Fujiki, R., Ito, S., Suzuki, E., Tanabe, M., Zhao, Y., Kimura, S., Fujiyama, S., *et al.* (2010). A histone chaperone, DEK, transcriptionally coactivates a nuclear receptor. Genes Dev *24*, 159-170.
Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A *103*, 1412-1417.
Schwabish, M.A., and Struhl, K. (2006). Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II. Mol Cell *22*, 415-422.
Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.
Sen, G.L., Reuter, J.A., Webster, D.E., Zhu, L., and Khavari, P.A. (2010). DNMT1 maintains progenitor function in self-renewing somatic tissue. Nature *463*, 563-567.
Shibahara, K., and Stillman, B. (1999). Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. Cell *96*, 575-585.
Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844-847.
Sims, R.J., 3rd, Chen, C.F., Santos-Rosa, H., Kouzarides, T., Patel, S.S., and Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. J Biol Chem *280*, 41789-41792.
Sims, R.J., 3rd, Millhouse, S., Chen, C.F., Lewis, B.A., Erdjument-Bromage, H., Tempst, P., Manley, J.L., and Reinberg, D. (2007). Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. Mol Cell *28*, 665-676.

Singhal, N., Graumann, J., Wu, G., Arauzo-Bravo, M.J., Han, D.W., Greber, B., Gentile, L., Mann, M., and Scholer, H.R. (2010). Chromatin-Remodeling Components of the BAF Complex Facilitate Reprogramming. Cell *141*, 943-955.
Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol *3*, Article3.
Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X., *et al.* (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol *29*, 68-72.
Sporn, J.C., Kustatscher, G., Hothorn, T., Collado, M., Serrano, M., Muley, T., Schnabel, P., and Ladurner, A.G. (2009). Histone macroH2A isoforms predict the risk of lung cancer recurrence. Oncogene *28*, 3423-3428.
Stargell, L.A., Bowen, J., Dadd, C.A., Dedon, P.C., Davis, M., Cook, R.G., Allis, C.D., and Gorovsky, M.A. (1993). Temporal and spatial association of histone H2A variant hv1 with transcriptionally competent chromatin during nuclear development in Tetrahymena thermophila. Genes Dev *7*, 2641-2651.
Suka, N., Luo, K., and Grunstein, M. (2002). Sir2p and Sas2p opposingly regulate acetylation of yeast histone H4 lysine16 and spreading of heterochromatin. Nat Genet *32*, 378-383.
Surani, M.A., Hayashi, K., and Hajkova, P. (2007). Genetic and epigenetic regulators of pluripotency. Cell *128*, 747-762.
Sutcliffe, E.L., Parish, I.A., He, Y.Q., Juelich, T., Tierney, M.L., Rangasamy, D., Milburn, P.J., Parish, C.R., Tremethick, D.J., and Rao, S. (2009). Dynamic histone variant exchange accompanies gene induction in T cells. Mol Cell Biol *29*, 1972-1986.
Suto, R.K., Clarkson, M.J., Tremethick, D.J., and Luger, K. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. Nat Struct Biol *7*, 1121-1124.
Szerlong, H., Hinata, K., Viswanathan, R., Erdjument-Bromage, H., Tempst, P., and Cairns, B.R. (2008). The HSA domain binds nuclear actin-related proteins to regulate chromatin-remodeling ATPases. Nat Struct Mol Biol *15*, 469-476.
Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell *116*, 51-61.
Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., *et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.
Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.
Talbert, P.B., and Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol *11*, 264-275.
Tazi, J., and Bird, A. (1990). Alternative chromatin structure at CpG islands. Cell *60*, 909-920.
Theise, N.D., and Wilmut, I. (2003). Cell plasticity: flexible arrangement. Nature *425*, 21.
Timinszky, G., Till, S., Hassa, P.O., Hothorn, M., Kustatscher, G., Nijmeijer, B., Colombelli, J., Altmeyer, M., Stelzer, E.H., Scheffzek, K., *et al.* (2009). A macrodomain-

containing histone rearranges chromatin upon sensing PARP1 activation. Nat Struct Mol Biol *16*, 923-929.
Tong, J.K., Hassig, C.A., Schnitzler, G.R., Kingston, R.E., and Schreiber, S.L. (1998). Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. Nature *395*, 917-921.
Torres-Padilla, M.E., Parfitt, D.E., Kouzarides, T., and Zernicka-Goetz, M. (2007). Histone arginine methylation regulates pluripotency in the early mouse embryo. Nature *445*, 214-218.
Trowbridge, J.J., and Orkin, S.H. (2010). DNA methylation in adult stem cells: New insights into self-renewal. Epigenetics : official journal of the DNA Methylation Society *5*.
Turner, B.M. (2000). Histone acetylation and an epigenetic code. Bioessays *22*, 836-845.
van den Berg, D.L., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010). An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell *6*, 369-381.
van der Heijden, G.W., Derijck, A.A., Posfai, E., Giele, M., Pelczar, P., Ramos, L., Wansink, D.G., van der Vlag, J., Peters, A.H., and de Boer, P. (2007). Chromosome-wide nucleosome replacement and H3.3 incorporation during mammalian meiotic sex chromosome inactivation. Nat Genet *39*, 251-258.
van Ingen, H., van Schaik, F.M., Wienk, H., Ballering, J., Rehmann, H., Dechesne, A.C., Kruijzer, J.A., Liskamp, R.M., Timmers, H.T., and Boelens, R. (2008). Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. Structure *16*, 1245-1256.
Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature *444*, 364-368.
Watanabe, S., and Peterson, C.L. (2010). The INO80 family of chromatin-remodeling enzymes: regulators of histone variant dynamics. Cold Spring Harb Symp Quant Biol *75*, 35-42.
Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet *39*, 457-466.
Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318-324.
Whitfield, M.L., Zheng, L.X., Baldwin, A., Ohta, T., Hurt, M.M., and Marzluff, W.F. (2000). Stem-loop binding protein, the protein that binds the 3' end of histone mRNA, is cell cycle regulated by both translational and posttranslational mechanisms. Mol Cell Biol *20*, 4188-4198.
Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343-348.
Wirbelauer, C., Bell, O., and Schubeler, D. (2005). Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias. Genes Dev *19*, 1761-1766.

Witt, O., Albig, W., and Doenecke, D. (1996). Testis-specific expression of a novel human H3 histone gene. Exp Cell Res *229*, 301-306.
Wong, L.H., McGhie, J.D., Sim, M., Anderson, M.A., Ahn, S., Hannan, R.D., George, A.J., Morgan, K.A., Mann, J.R., and Choo, K.H. (2010). ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. Genome Res *20*, 351-360.
Wu, C., and Gilbert, W. (1981). Tissue-specific exposure of chromatin structure at the 5' terminus of the rat preproinsulin II gene. Proc Natl Acad Sci U S A *78*, 1577-1580.
Wu, H., D'Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011a). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev *25*, 679-684.
Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Eve Sun, Y., and Zhang, Y. (2011b). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature.
Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011c). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389-393.
Wu, H., Tao, J., and Sun, Y.E. (2012). Regulation and function of mammalian DNA methylation patterns: a genomic perspective. Briefings in functional genomics.
Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. Nat Rev Mol Cell Biol *11*, 607-620.
Yap, K.L., and Zhou, M.M. (2006). Structure and function of protein modules in chromatin biology. Results and problems in cell differentiation *41*, 1-23.
Young, R.A. (2011). Control of the embryonic stem cell state. Cell *144*, 940-954.
Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res *21*, 1273-1283.
Zhang, R., Liu, S.T., Chen, W., Bonner, M., Pehrson, J., Yen, T.J., and Adams, P.D. (2007). HP1 proteins are essential for a dynamic nuclear response that rescues the function of perturbed heterochromatin in primary human cells. Mol Cell Biol *27*, 949-962.
Zhang, R., Poustovoitov, M.V., Ye, X., Santos, H.A., Chen, W., Daganzo, S.M., Erzberger, J.P., Serebriiskii, I.G., Canutescu, A.A., Dunbrack, R.L., *et al.* (2005). Formation of MacroH2A-containing senescence-associated heterochromatin foci and senescence driven by ASF1a and HIRA. Dev Cell *8*, 19-30.
Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev *13*, 1924-1935.
Zhou, H., Xia, X.G., and Xu, Z. (2005). An RNA polymerase II construct synthesizes short-hairpin RNA with a quantitative indicator and mediates highly efficient RNAi. Nucleic Acids Res *33*, e62.
Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J., and Melton, D.A. (2008). In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. Nature *455*, 627-632.

Zhu, D., Fang, J., Li, Y., and Zhang, J. (2009). Mbd3, a component of NuRD/Mi-2 complex, helps maintain pluripotency of mouse embryonic stem cells by repressing trophectoderm differentiation. PLoS ONE *4*, e7684.
Zofall, M., Fischer, T., Zhang, K., Zhou, M., Cui, B., Veenstra, T.D., and Grewal, S.I. (2009). Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs. Nature *461*, 419-422.